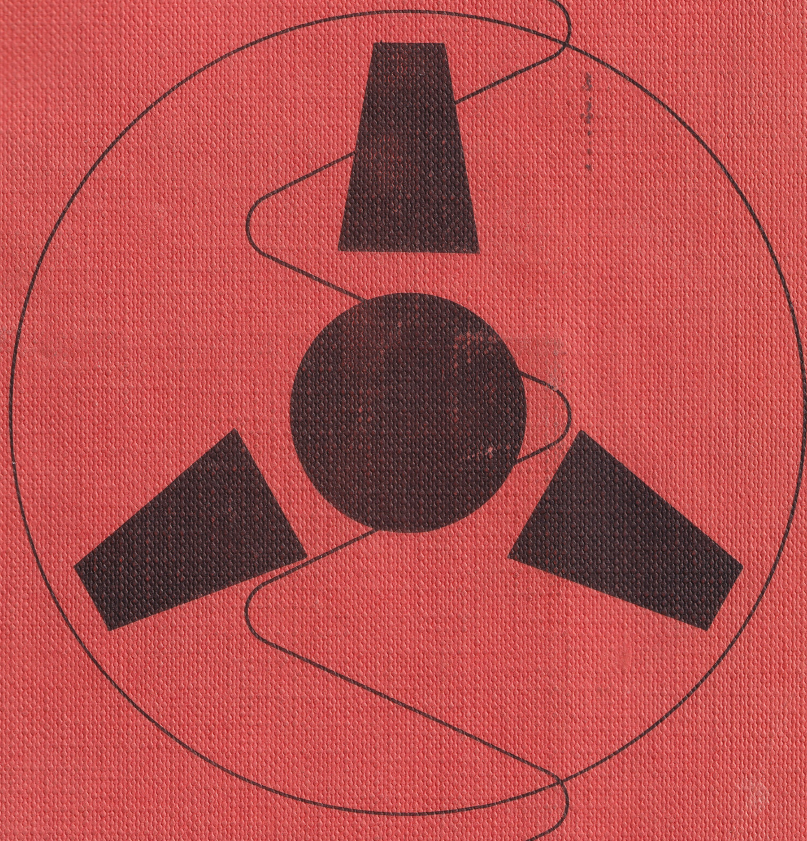


CHARACTER RECOGNITION



THE BRITISH
COMPUTER SOCIETY



PRODUCED BY UNEOPRINT
set on electric keyboards
photo-reproduced and printed offset

at The Gresham Press

UNWIN BROTHERS LIMITED

Old Woking Surrey England

(U 1468)

FOREWORD

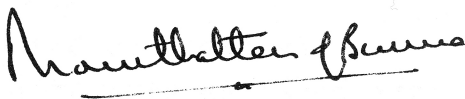
The British Computer Society has as one of its principal aims the dissemination of information to promote the wider and better use of computing techniques. It publishes The Computer Bulletin and The Computer Journal regularly. There is, however a need for works on specific aspects and the first of these was the Data Transmission Handbook 1964, which proved to be extremely successful.

The Character Recognition Handbook 1967 is the second in what is hoped will be a series on various techniques.

Character Recognition is something of a breakthrough in the problem area of data collection for computer processing and is used in association with new document handling methods. Its application helps to solve the cost, accuracy and timescale problems about which computer users have rightly been concerned.

There has been good progress in the U.K. in the use of magnetic ink character recognition, particularly in Banking, but the use of Optical Character Recognition is not yet as widespread here as it is in other parts of the world, particularly in the U.S.A. The number of applications is growing but much has to be done to assist in the use of computing techniques so important in the national interest.

The Society wishes to express its thanks to the equipment and services suppliers who have supported the Handbook by their contributions, and particularly to its own Character Recognition and Document Handling Group whose members have devoted so much time and enthusiasm to making this publication possible.

A handwritten signature in dark ink, which appears to read "Mervyn Bates". The signature is written in a cursive style and is positioned above a horizontal line.

President, British Computer Society

**BRITISH COMPUTER SOCIETY
DOCUMENT HANDLING & CHARACTER
RECOGNITION COMMITTEE 1966**

Chairman: Mr. F. H. SHARMAN

Secretary: Mr. A. E. HENDERSON

Messrs. C. W. BLAXTER

R. G. FIDDES

R. F. GOODMAN

W. A. GUNN

T. E. HARRIES

J. R. PARKS

S. R. C. POULTER

H. W. ROWBOTTOM

S. W. E. WILKINSON

The foreword to this handbook which appears on the previous page has been typed on an Olivetti machine fitted with one of the proposed international standard type faces for optical recognition, OCR Class B size 2.

In order to blend with the rest of this book the size of letters has been photographically reduced to 1 : 1.33 of its true size. This sample, also in OCR Class B size 2, is full size.

CONTENTS

Foreword by	Admiral of the Fleet The Earl Mountbatten of Burma K.G., P.C., G.C.B., O.M., D.S.O., F.R.S., D.Sc., M.I.E.R.E., M.I.E.E., President and Chairman of Council, The British Computer Society.	
Section 1	Introduction	3
Section 2	General—paper and printing characteristics	7
	Physical characteristics of the paper as they may affect the reading equipment	9
	Paper requirements for printing	11
	Printing considerations	13
	Effects of handling on paper and print	22
Section 3	Paper/print standards and fonts.	25
	Progress in font design to date	27
	The case for standardisation	28
	Design of character reading fonts	28
	Placing of characters to meet machine needs	30
	Print quality considerations	30
	Paper considerations	32
	Progress to date in standardisation	33
	Progress envisaged	36
Section 4	Document handling and transportation	37
	Introduction	39
	The basic requirements of a paper transport	39
	Outline of an automatic document handling system	40
	Speeds of operation	45
	Static Electricity	46
	Continuous stationery readers	46
	Maintenance	47
Section 5	Techniques of document scanning and character recognition	49
	Introduction	51
	Black box model of a pattern recognition machine	51
	Classes of character recognition systems	52
	Mark reading	52

	Magnetic ink character recognition	53
	Optical character recognition	55
	Output	70
	Future developments and prospects	70
	Further departures	73
Section 6	Proprietary equipment and services	75
	Automatic Input Systems Ltd.	77
	Bureauprep Ltd.	82
	Control Data Ltd.	84
	Crosfield Electronics Ltd.	90
	De La Rue Bull Ltd.	96
	Document Transports Ltd.	100
	English Electric Leo Marconi Computers Ltd.	102
	Farrington (Adrema Ltd).	110
	Honeywell Controls Ltd.	116
	I.B.M. United Kingdom Ltd.	118
	International Computers & Tabulators Ltd.	124
	Lamson Paragon Ltd.	130
	National Cash Register Company Ltd.	132
	British Olivetti Ltd.	140
	Recognition Equipment Ltd.	142
	George Waterstons & Sons Ltd.	148
Section 7	Cost Comparisons	149
	Keypunching, cards or paper tape	153
	Optical reading and writing data to magnetic tape	154
	Turnaround system	154
	Typing and optical reading	155
	Comparison of costs of keypunching cards and typing and optical reading	156
	Appendix	158
Section 8	Systems techniques—some practical applications	163
	Westminster Bank Ltd.	167
	Bankgiro, Sweden	174
	West Midlands Gas Board	179
	Legal and General Assurance Society Ltd.	182
	Montague Burton Ltd.	188
	Mercantile Credit Company Ltd.	192

SECTION 1

Introduction

INTRODUCTION

This Handbook is as much concerned with Document Handling, and with Paper and Print, as with Character Recognition—the latter has little commercial significance without the others—and there has been considerable difficulty in deciding the scope of the publication in certain areas. It was found, in fact, that some of the boundaries were as ill-definable as those of the poor-quality print characters that long-suffering reading machines are expected to recognise!

As far as Recognition and Handling equipment is concerned, it was eventually decided that Mark Reading by machine should be included, except where associated with punched card self-punching; and that the machine reading of document perforations should be excluded, except where the same perforations were disposed in the form of alphanumeric characters readable to the human eye.

Significance in Data Processing

There is ever-increasing emphasis being given to the problems of marshalling data external to the central processing function as such; indeed, the techniques and practice of Data Handling are assuming greater significance than actual main computing or file processing in many applications. The computer itself is assuming a greater role in this very field—whereby the processing function can be offered more rapidly, accurately and cheaply to more people—perhaps remotely—and in a wider range of application.

Within the concept of data handling, the field of Data Collection stands out as an area deservedly attracting a great deal of commercial attention; and anyone familiar with the implications of deriving accurate machineable data by preparation of punched cards or paper tape will know the substantial costs involved. These costs can be patently out of line with those applicable to the processing equipment to which the data is fed.

What can be done then to redress the balance? The obvious breakthrough is in dispensing with human transcription and a step forward can be taken by automatically deriving machineable data as a by-product of the transaction or event; this data may be transmitted over distances. Alternatively, original documents can be made machineable—and it is to this end that the techniques of Character Recognition and Document Handling are directed.

Such documents may have their data entirely computer-printed, for 're-entry' or 'turnaround' back into the system, subsequent to the transaction for which the document was created; e.g. utility bills or insurance premium renewals. In 'open-ended' applications, documents are prepared by printing press with pre-determinable data, and completed by keyboard entry when the transaction is over, e.g. bank cheques. Again, data to be read may be entirely keyboard-entered, and the latter process is not so far removed from creating punched cards or paper tape, except that no additional medium is required and the operation is

more akin to standard typewriter practice. Verification is the one headache of keying printed data for character reading, but encoding/inscribing equipments have been produced permitting such a function by various two-operator procedures and machine check-digit facilities.

Paper and Print

No document input and handling system based on character or mark reading can properly succeed unless paper and print are maintained to the standards laid down by the system designers and suppliers. Exact-ing standards are set for optimum system performances, but some relaxation from these may be tenable provided the consequent degradation of performance is within system viability limits. One of the biggest problems for the designer, in the current state of the art, is the quantitative assessment of the performance of a given system in relation to paper and print parameters; such information is as vital for improving or optimising the system as it is for estimating acceptable relaxation criteria.

Performance can be measured by the factors of speed of data throughput and by data reject and error rates. There is normally an inverse relationship between character reject and error rates—one may be traded for the other so to speak—and the right balance must be struck at the final design stages. A document may of course be rejected for paper handling rather than for recognition reasons.

Section 2 of the Handbook describes the characteristics of paper and print, without reference to standards for handling or recognition—whether actual or proposed. Paper properties likely to influence handling, printing, and character recognition are dealt with; and print parameters relevant to character reading are described as manifested in various printing processes and by various printing mechanisms. A final but important note is added about the effects of machine and manual handling on paper and print—these are vital considerations in any practical application!

Part of Section 3 deals with the paper qualities required specifically for character reading and, in the same context, with the questions of print quality and placement on the document.

The Type Font Controversy

Section 3 largely covers the issue of fonts for Magnetic Ink (MICR) and Optical (OCR) Character Reading, and the degree of standardisation achieved or anticipated.

In MICR, the situation leading to the adoption of E 13 B characters for Banking purposes in the USA and UK etc. is explained, and the current position in Europe with CMC7 is set out. Argument has raged on the relative merits of the two fonts, but recent comparative evidence favours CMC7 in many respects. The very substantial investment made by Bankers in E 13 B—certainly with very good results—and by their suppliers in reading, encoding and printing equipments would almost certainly deter them from making a change to CMC7 or to any other font without there being some overwhelming financial as well as technical advantage.

As regards OCR, the picture is much less clear, although several manufacturers' 'domestic' fonts have found wide application, mainly in the

USA. Recently, the USA have largely backed a class A highly stylised font, where many character features have been introduced to facilitate automatic reading, while ECMA in Europe have developed and backed a class B font, which is equally stylised but is aesthetically acceptable as well. It is expected that both will become standards approved by the International Standards Organisation (ISO), but the argument that class B characters can suffer greater degradation than class A, and still be readable, is finding an increasing following. The relative cost of systems based on the two fonts must, of course, be considered; certainly a class B reader will be more costly than its class A counterpart in a numeric only system. If the character repertoire is large the difference is not significant. No attempt is made to compare and contrast these two character sets, neither of which is yet in substantial practical use.

Techniques in Handling and Reading

In the majority of applications, documents for machine reading are of paper substance; and the present stage reached in the field of machine reading is as much due to developments in paper handling equipment as it is to progress in recognition systems. Documents must be fed from an input stack, transported at high speed—and in accurate register—through a reading stage and then placed in one or more output stacks for removal from the system or re-passage through it. Feeding must usually be automatic during a particular pass but, in multiple pass applications such as cheque sorting, there is manual handling between successive passages of the same documents.

Section 4 is devoted to design factors in paper handling devices; the requirements of document loading, feeding, transportation and stacking are explored separately, and in practical combination. It must be remembered that the paper in question has very rarely been 'untouched by hand', and the mechanisms already available are remarkably tolerant of all conceivable forms of spoilage. Machines designed for re-pass purposes are equally remarkable in the absence of spoilage they themselves impose.

Section 5 covers Character/Mark Scanning and Recognition techniques as such and it is important to note that the total Reading process has these two distinct elements. In fact, these elements are each further divided, and the resulting distinguishable stages are described in relation to Mark, Magnetic Ink and Optical Character Reading.

On the question of choice of MICR or OCR, it is now clear that MICR practice will not extend to any significant degree beyond banking and similar applications. Only where security is at a premium, and machine readability with visual obliteration (overstamping, etc.) is essential, can the exacting MICR printing and encoding standards be justified economically.

For the general run of commercial applications, indifferent printing standards—albeit with standard type fonts—will have to some extent to be tolerated. Computer output printing must certainly be machine legible, for example, and MICR printing speeds have proved too slow for the substantial document producing loads found in major application, such as utility billing. For commerce, then OCR will largely win the day—in many cases in conjunction with Mark Reading.

Likely future developments are summarised and the possibilities for handwriting recognition, for example, are put firmly in perspective.

Equipment and Facilities. What is Available ?

No comprehensive Handbook would be complete without information on what machines and services are offered by manufacturers and suppliers. Section 6 is concerned entirely with these matters, and a large number of companies have taken space to set out what it is they have to offer the user from the wide technology which has grown up around the basic character reading art. Many reading and encoding equipment ranges are described, as are a variety of facilities in the area of 'paper and print'.

Economics, Systems Techniques and Some Live Applications

Section 7 is an attempt to justify character reading on relative cost grounds, in comparison with other forms of data input and collection. The task is no easy one, but the issue has not been 'ducked' !

In support of the case 'in favour', a representative set of actual working systems is described in Section 8, following a preamble on systems techniques generally. Mark Reading and MICR are covered in addition to OCR, and important matters such as reject and throughput rates are quoted, wherever it has been possible.

Usage to Date

In MICR, the investment in sorting/reading and encoding equipment is already enormous. The number of cheque sorters, for example, installed worldwide for commercial banking purposes is estimated in late 1966 to be into the four figure bracket, with the USA and Western Europe accounting for virtually the complete total. The UK and Commonwealth banks have followed the USA lead with the E.13 B font, whereas Continental Europe has already largely backed CMC7.

OCR has not, as yet, made the same sort of progress, but substantial increases in usage are expected in the remaining part of this decade. Apart from commerce generally, there are signs that MICR may be ousted by OCR in areas so far regarded as the almost certain prerogative of the former; the European Post-giros provide a very substantial and largely uncommitted sector in this respect. Apart from Mark Reading the total of OCR installations of all types in late 1966 is in excess of five hundred, of which Western Europe accounts for under 10%. Final decisions on ISO fonts will contribute in forcing the pace, particularly in 'open-ended' systems, but it must be remembered that some applications, particularly of the 'turnaround' type, can be OCR-based with any style of font; the success of OCR to date is certainly some reflection of this fact.

References

Achilles, C. 'Vergleich der Schriftsysteme E13B und CMC7 in einem praktischen Versuchsbetrieb der Deutschen Bundespost' Bankbetrieb 3:4; 1963.

See also summary of comparative tests in:-
W. Dietrich, 'Optical Character Readers for Automatic Handling in Banking Applications' Electrical Communication 40:3; 1965.

SECTION 2

General

Paper and Printing Characteristics

GENERAL—PAPER AND PRINTING CHARACTERISTICS

PHYSICAL CHARACTERISTICS OF THE PAPER AS THEY MAY AFFECT THE READING EQUIPMENT

This section outlines in general terms the properties of paper and their possible effects on the performance of a Reader. It does not presuppose any particular method of handling or specify any type of Reader. Thus not all the properties referred to will apply to every system. However, the more fully the user of a Reader is able to specify the paper requirements, the more likely is the papermaker to be able to meet them.

Paper properties are considered firstly as they may affect the transporting and positioning of the document and secondly as reading may be affected.

No attempt is made to suggest standards, since work on standardization is being done by several other bodies. The aim here is to indicate the significance of the properties and, where possible, to give methods of measurement.

Handling

Dimensions. Maximum and minimum thickness or caliper of paper which can be accepted may be determined in practice by the range of stiffness which can be tolerated.

Paper may vary from sheet to sheet by $\pm 10\%$ in thickness, but such variations are likely to influence consistency of results during printing rather than at the document handling stage

Stiffness. This can be measured directly^{1,2} and will be related to substance and grain direction.

Substance. The weight per unit area of a paper.

When paper is bought in quantity it is priced by weight and substance is conveniently expressed as the weight in pounds of a ream of, say, 500 sheets of a certain size, say 20×30 in. Otherwise, substance may be expressed in g/m^2 . Multiplication factors to assist rapid conversion from one system to the other are available.³ The substance will to some extent influence opacity, requirements for which may set the lower limits on substance.

Flatness. This could affect transport and positioning alike and, since documents will often have been handled by the public beforehand, equipment will probably be designed to accommodate some departure from flatness provided that the document is not too stiff.

Grain direction. Paper is made on a continuously moving fine wire mesh and the fibres are distributed somewhat directionally. Paper properties thus differ in the grain or (paper making) machine direction from those in the (cross) direction at right angles.

Owing to the directional disposition of the fibres, stiffness is generally greater in the grain direction and the rigidity of a document will thus depend upon whether its longer dimension runs in the grain or cross directions of the paper. How the document is to be presented to the Reader may determine the choice of grain direction.

Friction. Depending upon the method used to transport the document, the coefficient of friction either between paper and paper or paper and some other surface may be important and can be measured.⁴ Printed matter other than that required for document processing (e.g. security backgrounds) can influence the surface friction of the finished document; so also can anti set-off powder which may be deposited during each printing sequence. (See Printing Considerations, p. 00)

Porosity. Although most papers will present no difficulty, if document control depends upon suction, the air porosity of the paper may be relevant and can be measured on a number of instruments.⁵

Moisture content. Control of documents may be adversely affected by the presence on them of electrostatic charges. These will be more troublesome the lower the moisture content of the paper and this will be largely determined by the relative humidity (r.h.) of the atmosphere in which the documents have been stored immediately prior to processing.

The paper maker can, within limits, adjust the moisture content of the paper to be in equilibrium with a controlled r.h. in the printing works.

This would be of little significance at the data processing stage however, unless the document spent its entire life in the same atmospheric conditions.

The main point is that storage conditions immediately prior to processing should not be too different from those during processing and the r.h. should not be too low (50%). An appreciable difference in r.h. between storage and reading conditions may result in a lack of flatness in the documents.⁶ For small documents, static is likely to be more troublesome than departures from flatness.

Tear strength. The ability of the document to resist general wear and tear could be assessed by some form of tear test; for example a test⁷ which measures the force required to extend an existing small tear would perhaps reflect the way in which prior mishandling might affect the passage of the document through a Reader.

Reading

Paper properties which affect reading are those which can interact with the image and these are most numerous in the case of optical character recognition.

Holes. The presence of holes in the paper within the area allocated to encoding is likely to cause trouble to any Reading system and, although some may occasionally result from deficiencies in manufacture, their likelihood is probably greater when the paper has become a document and holes from pins, staples or punches may appear.

Specks. Particles occurring in the paper surface can be equally troublesome.⁸ In addition to their number and size, their reflectance (for OCR), magnetic properties (for MICR) or perhaps electrical conductivity may have to be controlled.

Brightness. Foremost among the optical properties affecting OCR is the average reflectance of the paper relative to a standard surface such as magnesium oxide or magnesium carbonate; a more durable intermediate reference surface such as a plastic chip or ceramic tile may be used for routine testing but it should be checked regularly for yellowing or other changes. Generally, the test paper should be backed by a small pile of similar sheets since most papers are not completely opaque.

There are standard methods of measuring brightness.⁹ Apart from geometrical features, an important consideration is the colour of the light used. This is generally related to visual considerations, the characteristics of the eye and the type of lighting most likely to be encountered in practice. In standardizing for OCR applications the light used in scanning and the spectral sensitivity of the photo-detectors should form the criteria.

For near-white papers the choice might not be critical although many, but not all, contain 'optical bleaches', fluorescent dyes absorbing in the near u.v. and emitting in the blue region of the spectrum. In some cases, coloured papers may have a sufficiently high brightness to be acceptable.

Opacity. As was indicated earlier, lack of opacity in a paper may reduce its effective brightness unless it is backed by a white surface during reading. If the backing is black, or a cavity, or a printed document—or if the document is printed on the back, opacity may be an important consideration.

It can be measured as printing opacity,¹⁰ the ratio of the brightness of a sheet of paper when backed by a black cavity to that when it is backed by a pile of similar sheets.

Gloss. This can be measured and specified if necessary.¹¹ In general, optical systems in Readers will be designed to exclude most of the specularly reflected light from the photo-detector.

*Linting or fluffing.*¹² The production of dust or fibres from either the surface or the cut edges of the paper may lead to deposits on optical surfaces in the Reader, producing slow changes in performance.

*Abrasiveness.*¹³ If a transparent plastic or glass window were used through which the document was to be read, scratching by abrasive particles in the surface might impair reading.

PAPER REQUIREMENTS FOR PRINTING

Smoothness. Printing for OCR applications is, initially at least, likely to be predominantly from such sources as typewriters and high speed printers. The most important paper property affecting the print quality will be surface smoothness.

Printing for applications in the field of MICR where requirements are more exacting has been associated more with letterpress and offset lithography. Again surface smoothness is important and several tests, most of which test the surface when it is under pressure, have been devised.¹⁴

Absorbency. This is an important property in that it affects the image as laid down, for instance its gloss, and also the likelihood of impairment immediately after printing with liquid inks due to smudging or to 'setting off' onto the back of the sheet laid on top in the delivery pile. Absorbency can also influence rubproofness of the finished print. Apart from very broad generalizations these aspects of paper absorbency cannot be considered without taking into account the properties of the printing ink. What is desirable is that there should be uniformity of absorbency over a sheet and between successive sheets.¹⁵ Uniformity of smoothness is also important and in both cases may be affected by the formation of the paper, the evenness of distribution of fibres in the sheet.

*Thickness.*¹⁶ The thickness of paper should not vary unduly over a sheet nor from sheet to sheet or print quality will suffer. Possible effects of water marks should be borne in mind.

*Fluffing or linting.*¹² The production of dust or fibres from either the surface or the cut edges of the paper may result in a deterioration in print quality, and, in the case of lithography, of the image on the plate.

Surface strength.^{17,18} As paper and type separate after the instant of printing, tack forces in the ink may be sufficient to tear fibres or other pieces away, leaving voids in the printed ink film. Debris remaining on the printing surface may affect succeeding prints. Satisfactory performance under given printing conditions requires that the surface strength of a paper should be adequate.

Moisture content. This must be considered in relation to the atmospheric conditions in the printing machine room for reasons similar to those outlined in the case of the Reader, (See p. 10) since the paper transport on the press can be similarly impeded by lack of flatness of the sheet or by the presence of static.

Furthermore, paper dimensions vary with moisture content and paper which has lost its flatness due to changes in moisture content can result in distortion of the printed image. Since, in commercial printing, small documents may be printed in groups on a large sheet, accurate register between successive printings, or between printing and guillotining, may be lost if the paper is exposed to varying temperature and humidity. Dimensional change is greater in the cross direction and some papers are more sensitive than others; stabilities can be compared and standardized if necessary.⁶ Where temperature and relative humidity in the machine room are controlled, paper may be ordered with a moisture content in equilibrium with the r.h. This should be kept in moisture-proof wrappers which should be removed in the machine room only after sufficient time has elapsed for the stack of paper to reach temperature equilibrium.

pH. Ink drying can be affected by acid conditions in lithographic papers and pH should not fall below 5.^{19,20}

PRINTING CONSIDERATIONS

The more obvious applications of OCR are likely to involve printing methods distinct from normal processes, and in some cases giving inferior print quality. Furthermore, as Readers become more sophisticated, the standard of minimum readable quality will probably eventually fall to a level below what is acceptable in normal commercial printing. On the other hand for MICR, quality tolerances are much closer; requirements were laid down by the Committee of London Clearing Bankers²¹ which were rather more exacting than those which apply to normal commercial printing. Guidance has been given on the practical problems of meeting these requirements.²²

There are a number of features of conventionally printed characters which may prejudice the performance of character recognition systems.

- (a) Variations in ink film thickness and blackness across a sheet.
- (b) Variations in stroke width owing to ink squash.
- (c) Non-uniformity of ink across a character or stroke, including voids.
- (d) Irregularity of stroke boundaries.
- (e) Deformation (embossment) of the paper by the printing surface.
- (f) Extraneous ink on the face of the sheet.
- (g) Extraneous ink on the back of the sheet.
- (h) Long term changes in stroke width owing to wear of the printing surface.
- (i) Sheet-to-sheet variation of (a), (c) and (d).
- (j) Presence of anti set-off powder deposited after printing.

Against this list of features, the chief characteristics of the three principal commercial printing processes may be considered. The processes are dealt with here in order of decreasing importance to character recognition applications under present conditions.

Letterpress

Printing is from a relief surface, Fig. 1(a), the non-image areas being too low to contact either the inking rollers or the paper during printing. Type is generally cast or moulded from matrices; it may be etched through a photo-resist.

The process is the one best suited to the incorporation of consecutive numbering, where printing is, in this case, from mechanically engraved discs.

Of the features listed above (b), (c), (e) and (h) are of particular concern.

The printed characters have a well defined edge, Fig. 2; the printed ink film is often uneven, the optical density of a stroke being higher towards the edges than in the middle owing to ink-squash. The prominence of this effect depends upon ink and paper properties and may change as the edges of the type characters become rounded with wear.

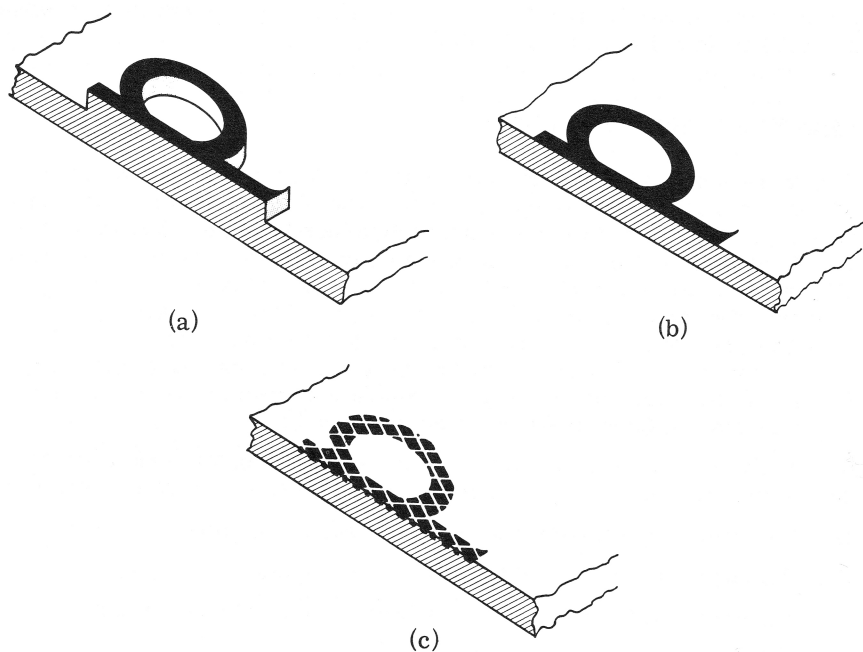


Fig.1 Diagrammatic representation of printing surface for (a) letterpress, (b) lithography and (c) photogravure.



Fig. 2 Example of letterpress printing

During printing, the paper is backed by a 'packing' of paper (and, perhaps, plastic or rubber) sheets which permits roughness of the paper surface and slight dimensional inaccuracies in the type height and the press to be taken up.

In general, the greater these deficiencies, the softer must be the packing and the greater will be the tendency for embossment to occur.

Offset lithography

Printing is from a virtually flat surface, Fig. 1(b); the non-image areas are water-receptive and an applied water film prevents the deposition of ink in those areas. The image is usually offset first onto a rubber-faced blanket and thence onto the paper.

The image is prepared photographically starting from letterpress type, letterpress prints, photographic masters, or typewritten matter, or can be typed directly onto a paper plate. Features (a), (f), (h) and (i) are of

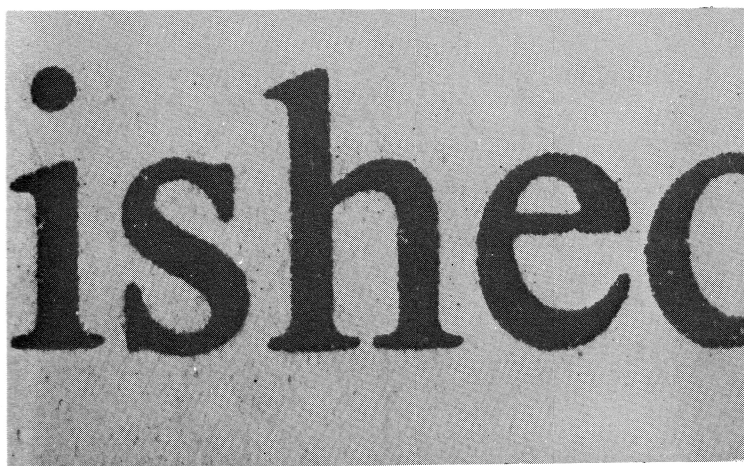


Fig. 3 Example of offset lithography printing

importance. The print generally has a less sharply defined edge, Fig. 3, but the ink distribution is more uniform as compared with letterpress.

In both letterpress and lithographic printing, 'setting-off', the transfer-ence of ink from a freshly printed sheet onto the back of the sheet laid on top, is frequently minimized by spraying a very small quantity of anti set-off powder (finely powdered limestone or starch) onto each sheet. As already noted, this could affect frictional properties of the paper surface and, possibly, optical properties of the print.

Photogravure

Printing is from a recessed surface, Fig. 1(c), the image consisting of small cells which contain the ink. The non-image area contains no cells and its surface is scraped free of ink by a doctor blade.

Features of importance are (c), (d) and (f). Type characters tend to have a serrated edge, owing to the cell structure, but this may become

indistinct owing to ink flow on the paper. Generally, the shape and dimensions of characters produced by this process are less easily controlled than those printed by letterpress and lithography.

The printing surface is prepared by photographic and chemical means and the high cost of production means that use of the process is generally confined to long printing runs. It is thus unlikely that photogravure printing, Fig. 4, will impinge much on either MICR or OCR. However, it

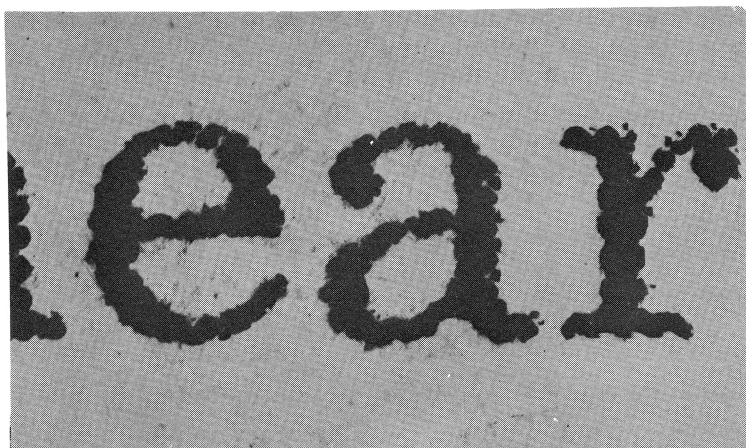


Fig. 4 Example of photogravure printing

is the process by which many women's magazines and newspaper colour supplements are printed and it is conceivable that such things as publication codes or dates on mail order coupons, for example, might create a demand for OCR processing.

Other processing

Other methods of printing and copying are generally well known but it is of interest to compare characteristics.

The illustrations show examples of print which may not be the best obtainable in any class but which serve to illustrate certain features.

Typewriters, embossed plate printers, encoders and high speed printers can all produce a certain amount of embossment. The sharpness of the image depends on the condition of the ribbon and whether it is an inked ribbon or a complete transfer type.

There are other characteristics peculiar to the different printers.

Typewriters. Uneven density due to variations in operator touch may exclude manual typewriters from some OCR uses. Depending on the condition of the typewriter, variations in spacing between characters and skew may result in unacceptable quality, and from this point of view typewriters printing from a 'head' or 'golf ball' may be more satisfactory.

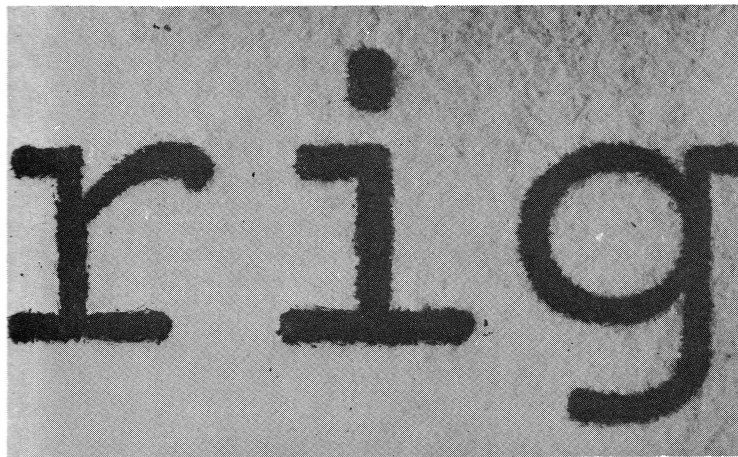


Fig. 5 Typewriter with total transfer ribbon

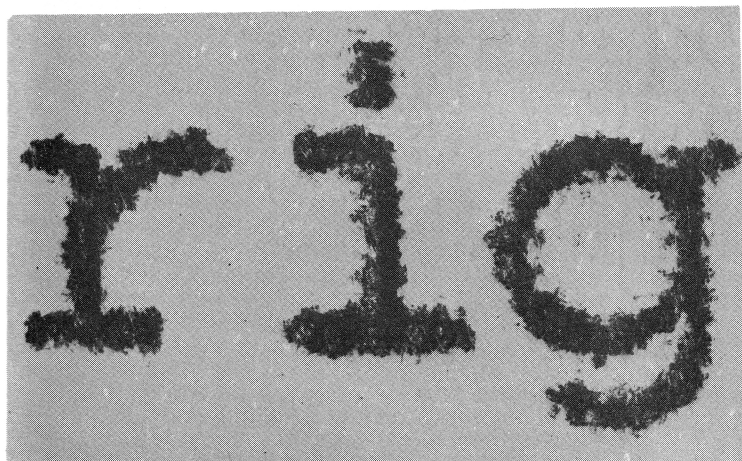


Fig. 6 Typewriter with inked ribbon

Embossed plate printers. Unlike the established applications such as addressing, systems likely to use OCR methods involve the use of 'positive' or 'right-reading' raised characters. This entails printing through the paper, with the result that the print quality suffers, and also that the paper may be thinner than would otherwise be desirable. Since the plates are carried about by the owner, there is a risk of accidental damage to characters.

Encoders. Fig. 8. This class of printer usually gives good quality print having been developed for encoding cheques for MICR systems where requirements are particularly exacting. The pressure is con-



Fig. 7 Embossed plate printing



Fig. 8 Encoder with total transfer ribbon

stant over a whole line. Results are influenced by the ribbon which should be of the total transfer type. 'Hard' ribbons tend to give irregular edges whilst 'soft' ribbons give some squash with a tendency for voids to occur in the centre of the stroke. (In general, both embossed plate printers and encoders have been designed to handle specific sizes of forms and to print in pre-determined positions.)

High speed line printers (for computer output) generally print from continuously moving characters the quality and alignment of print depending upon accurate timing and short contact periods.

Line printing from a drum or barrel. Fig. 9. The entire repertoire of characters occupies one ring around the barrel, the maximum number of characters in a line determining the total number of rings. The barrel rotates in a direction perpendicular to the lines of print at a constant speed which may be as high as 1500 r.p.m., and hammers press the paper and ribbon against the required character at the correct instant. There is inevitably some thickening of horizontal strokes owing to the finite contact time although this may be only a few microseconds. Horizontal spacing of the characters is fixed by the rings on the drum. Random timing errors can result in variations in horizontal alignment of printed characters. Overall drifts in timing caused, for instance, by 'warming-up' may result in weak tops or bottoms of the printed characters unless adjustments are made.

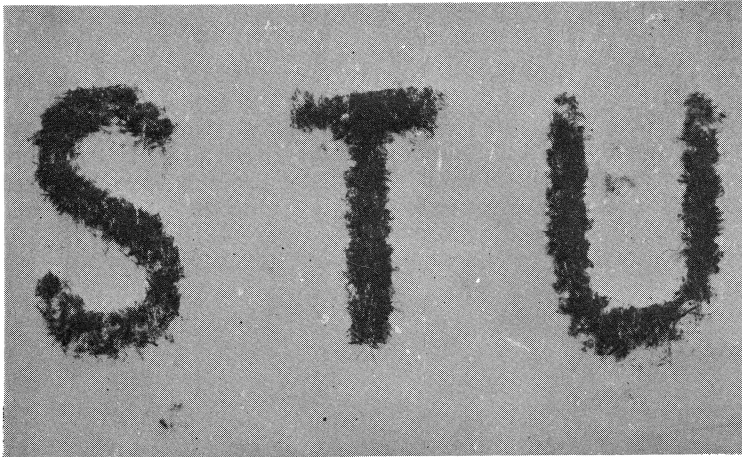


Fig. 9 High speed drum printer

Excessive stretch in the paper may lead to curvature of lines of print across the page due to stress imparted to the paper through the ribbon by drum movement. Similarly, when printing covers only a small part of a large document, forces occasioned by rapid repositioning of the paper may necessitate a brief delay in printing to enable the paper to recover if curving of lines is to be avoided.

Line printing from a chain. Fig. 10. Type characters are mounted on a continuous chain moving in the direction of the lines of print. Provided that the chain does not 'snake', horizontal alignment is good, but timing may affect spacing between letters and contact time will affect thickening of vertical strokes.

It should be noted that although relief printing is employed, all these methods differ from letterpress not only in the use of some form of ribbon but, more fundamentally, in the fact that they are generally only intended to produce one copy.

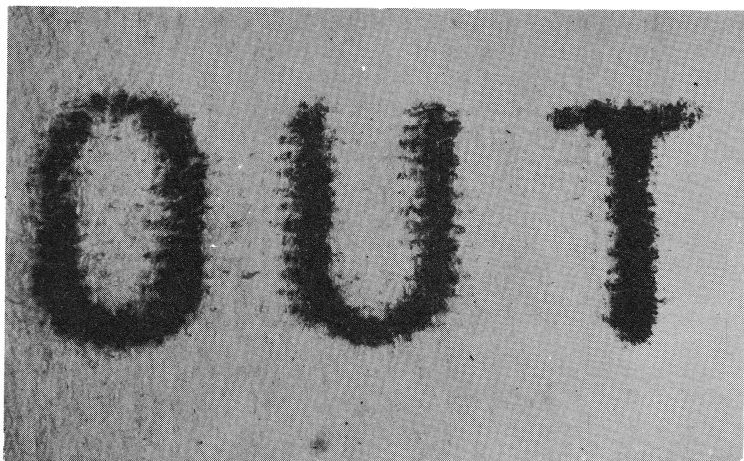


Fig. 10 High speed chain printer

Where a number of copies are required, carbon paper, Fig. 11, stencils, Fig. 12, xerography, Fig. 13, or offset lithography may be used depending on the quantity and quality required. These produce very different results, the last mentioned process usually giving the best definition.

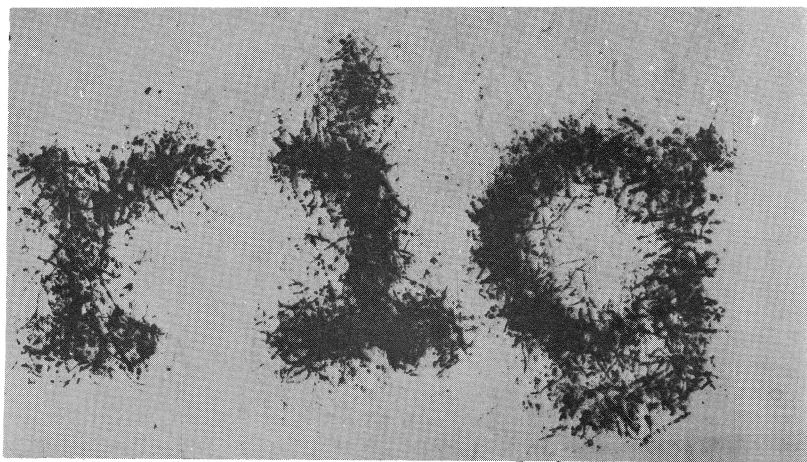


Fig. 11 Second carbon copy

Line printing from cathode-ray tube display. In order to overcome some of the problems of high-speed printing by mechanical means, the Xeronic method produces a line of characters on a CRT display which is recorded xerographically on a moving web of plain paper, the CRT image being moved synchronously. Printing speeds of up to 4,700 characters/sec. are possible. Developments include superimposition on the type image of a form image of rules and headings. Simultaneous display on a second CRT for producing a copy document is also possible. The image quality is typical of xerographic printing, with the possibility of distortion in character dimensions at the edges of the CRT face also.

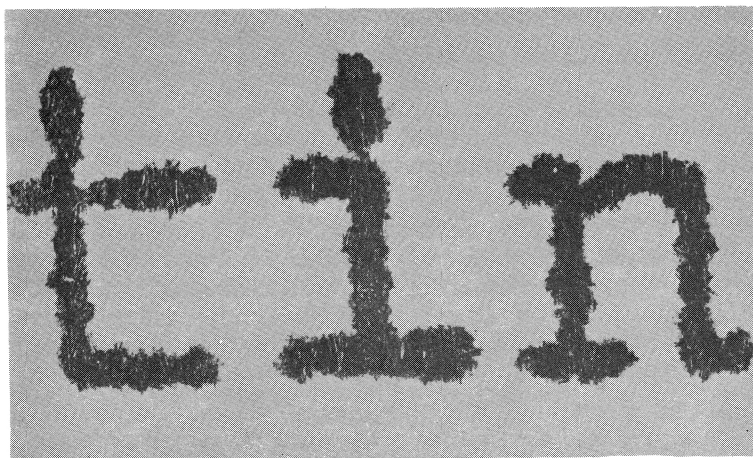


Fig. 12 Stencil

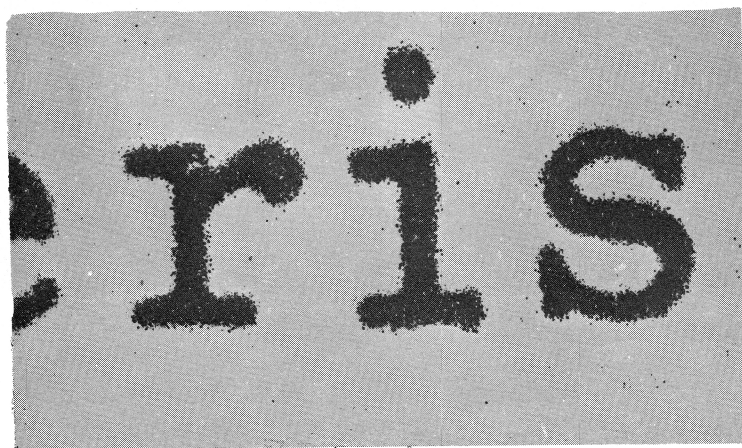


Fig. 13 Xerography

Allied to this are several systems designed for photo-typesetting (i.e. producing type matter on photographic film or paper of quality suitable for reproduction in book or publication printing), the input being from tape or computer output. These should be mentioned, although they are somewhat elaborate to be used simply as a means of getting high print quality in computer outputs for OCR purposes. They include the Hell Digiset (up to 1,000 ch/sec.), an RCA system in which 5,000 ch/sec. are envisaged and also the Lumizip 900, with speeds of up to 600 ch/sec., in which the characters originate from an optical type font on glass.

Measurement of print quality

Sharpness of printing and the accuracy of stroke widths can be assessed by throwing an enlarged image of a character onto a screen bearing a reference character or graticule, and observing discrepancies. If more quantitative methods are required it is important that the quantities measured are strictly relevant and because of this, instruments to measure quality are likely to be developed in conjunction with reading equipment. There are two commercially available examples of such instruments.

The 'Micrex'²³ is designed to monitor MICR characters where the visual pattern is not necessarily a reliable guide to the magnetic signal. The character wave-form is displayed on the screen of a cathode ray tube where the shape and amplitude can be checked on a graticule.

Equipment for monitoring OCR characters²⁴ also provides a CRT display, in this case of the magnified outline of a character as 'seen' by a 0.2 mm scanning spot. The observed density gradient at a stroke boundary and the severity of voids or of extraneous inked areas are obviously functions of scanning spot size and thus Monitor and Reader should have comparable optical characteristics. The displayed boundary can be selected for any brightness level between the white of magnesium oxide and absolute black. Fig. 14 illustrates the effect of varying this

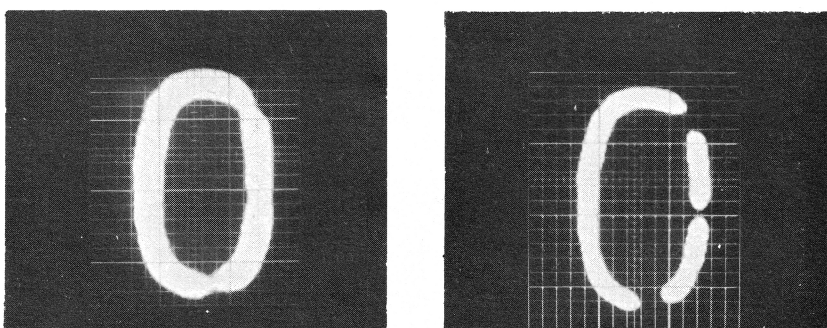


Fig. 14 Character O with voids at PCS levels of (left) 0.51 and (right) 0.70.

level from 0.51 to 0.70 on the PCS scale where white paper gives 0 and absolute black gives 1.00.

EFFECTS OF HANDLING ON PAPER AND PRINT

Taking the word literally, handling of paper can result in curl of the sheet, creasing, local distortions through uptake of moisture, and discoloration and scuffing of the printed characters; none of these effects can be dealt with very specifically. The best way to minimize the effects of handling would be to select a fairly stiff, smooth, not too glossy paper, of good dimensional stability and with high tear strength.

With the exception of news ink on newsprint, Graphic Arts printing in general gives an image of good rubproofness. Office and document

printing should also give satisfactory results in most cases except for carbon copies, which will probably have been excluded from character recognition applications on other grounds. As regards ink, if it is based on carbon black, light fastness should be no problem.

References

'Proposed procedures for the testing of paper and boards', are published by the British Paper & Board Manufacturers' Association of Great Britain, Plough Place, Fetter Lane, London, E. C. 4.

British Standards are published by the British Standards Institution, 2 Park Street, London, W. 1.

TAPPI Test methods are published by the Technical Association of the the Pulp & Paper Industry, 360 Lexington Avenue, New York 17, N. Y.

- (1) TAPPI-T451m, 'Rigidity, stiffness and softness of paper'.
- (2) TAPPI-T489m, 'Stiffness of paper and board'.
- (3) B. S. 730, 'Specification for papers and boards'.
- (4) B. S. 2782, Pt. 3, Method 311A, 'Coefficient of friction of polythene film'.
- (5) B. S. 2925, 'Methods for determining the air permeability and air resistance of paper'.
- (6) Pritchard, Evelyn J., 'Some effects of relative humidity and temperature on paper and inks', Brit. Ink Mkr., (1963), 6 (1) 38.
- (7) TAPPI-T414ts, 'Internal tearing resistance of paper'.
- (8) TAPPI-T437ts, 'Dirt in paper and paperboard'.
- (9) TAPPI-T452m, 'Brightness of paper and boards'.
- (10) B. S. 2923, 'Printing opacity of paper'.
- (11) Harrison, V. G. W. and Poulter, S. R. C., 'Gloss measurement of high-gloss papers', Research (1954) 7 (4) 128.
- (12) Proposed Procedure No. 8, 'Estimation of fluffing tendency by the PATRA fluff tester'.
- (13) Bekk, J., 'What the printing process demands of paper'. Proc. Tech. Sect. P. M. A. (1935) XVI, 9.
- (14) Peacock, E. W., 'Some aspects of printability of paper and board', Brit. Ink Mkr., (1961) 3 (4) 21.
- (15) Proposed Procedure No. 1, 'Ink absorbency of board by use of K and N ink'.
- (16) B. S. 3983, 'Methods for the determination of the thickness and bulk of paper'.
- (17) TAPPI-T499su, 'Surface strength of paper (IGT tester)'.
- (18) Proposed Procedure No. 12, 'Picking strength by the Dennison wax method'.
- (19) Coupe, R. R., 'Acid papers and ink drying', Penrose Annual (1954) 48, 113.

- (20) B. S. 2924, 'pH values of aqueous extracts of paper'.
- (21) 'Requirements for automatic cheque processing', The Committee of London Clearing Bankers, 10 Lombard Street, London, E. C. 3.
- (22) Boxall, L. E. and Clayden, D. O., 'Magnetic ink printing', PATRA Principles in Practice No. 2. The Printing, Packaging & Allied Trades Research Association, Leatherhead, Surrey.
- (23) The Micrex magnetic ink printing tester, Crosfield Electronics Ltd., 81 Fortess Road, London, N. W. 5.
- (24) Optical Print Quality Meter, ICT, Cavendish Road, Stevenage, Herts.

SECTION 3

Paper/Print Standards and Fonts

PAPER/PRINT STANDARDS AND FONTS

Progress in font design to date

It has already been explained in the Introduction to this book how automatic reading of humanly legible characters can simplify the problem of data handling and help to reduce costs. These advantages can be obtained by having the visually acceptable and machine readable characters imprinted in the same operation upon a document and ideally they should be combined into the same script. The machine reading process is, as elaborated further in Section 5, essentially a matter of detection, with pulses created within the machine in recognisable patterns that identify the character under the reading head. It is clear therefore that the code which in some manner or other is contained within the character on the document must fulfil such standards as are set by the reading machine.

This constraint upon the character means that the enormous variety of hand written scripts represents probably the extreme of difficulty for the designer of a reading system and that even any mixing of machine print either in quality or style was beyond the bounds of practical possibility in the early attempts to achieve automatic character recognition. Rather, the first approaches were to devise a limited range of stylised numerals and a reading machine which were mutually compatible. Inevitably each manufacturer of a reading machine produced his own font and the wide range of fonts in existence may be appreciated from the small selection illustrated in Fig. 15.

SOME TYPE FONTS FOR OPTICAL CHARACTER RECOGNITION

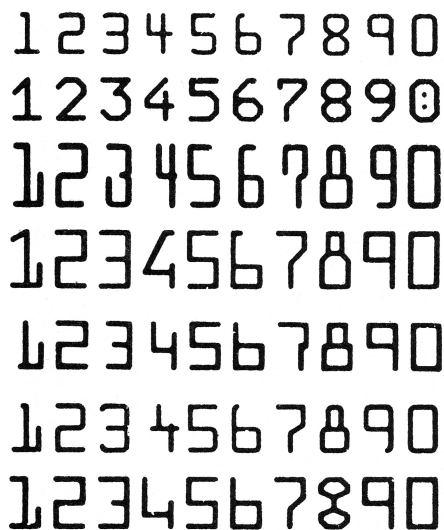


Fig. 15

Unfortunately, the economic and practical dictates of particular reading techniques and the incorporation of special features (for instance to provide a degree of indifference to mutilation of the character code-line(s) by extraneous marks or damage to the document) led to many of the fonts departing widely from the accustomed styles of print. In some situations this may be acceptable but unfamiliar fonts are unlikely to be acceptable for ordinary reading by the general public.

As development in this field progressed the repertoire of characters within specific fonts was extended to include alphabetic as well as numeric characters and symbols. At the same time pronounced departure from conventional styles of print has become less necessary as the tolerance of the reading device has been improved and so a measure of uniformity has come about. Standardisation upon fewer fonts has therefore appeared to have come within reach.

The case for standardisation

In these circumstances it can reasonably be expected that the interchangeability of documents between ADP Systems and across company boundaries, will be greatly facilitated. This should bring about among the general public greater familiarity with machine-readable fonts and thereby increase their acceptability. There is also likely to be simplification of the demands upon stationery printers, and designers of both printing machines, large and small, and reading devices. Furthermore, changeover to character reading systems would be less likely to demand replacement of typewriters, tally roll printers etc., if type-faces which are acceptable to reading devices were being used already. The total effect should be to allow reduction of the costs of data processing at large, and this argument is developed further in Section 7.

Design of character reading fonts

It is worth recalling that human legibility depends upon the overall shape of the character whereas the ability of a machine to read a character depends upon the presence or absence of contrasts in particular positions. Significant improvement in reading accuracy by machine can be obtained if the character shapes are as distinguishable as possible in terms of the capability of the reading device.

Typographical designers have displayed boundless ingenuity in the design of characters to meet specific practical or aesthetic requirements. This has given rise to designs of characters employing variations in stroke widths to provide variety in appearance, variable pitch between characters to take up the differences in apparent bulk of the character shape and various decorative features, such as serifs. Visual effectiveness may also be aided by mixtures of fonts—it is popular among writers to indulge in italics for instance—and where would we be without capitals to start our sentences?

From the point of view of an automatic reading device most embellishments such as serifs are superfluous, if not positively confusing, and at the present stage of development of reading devices uniformity both in stroke widths and spacing of characters are basic requirements. In

addition, the font designer is usually required to accommodate the idiosyncrasies of the reading device by incorporating characteristics which from the human reader's standpoint were better absent. For these fundamental reasons it is still not generally possible to employ the commonplace type-fonts for low-cost automatic reading systems.

There is an obvious relationship between the number of characters in a repertoire and the number of recognition points which have to be built into the character shapes so that each may be as distinguishable as possible. The earlier character fonts were largely confined to numerals together with a few control symbols which, for example, can be used to prompt the reading of specific field. With some reading methods it is desirable that every character includes in its shape a feature in a specific position which acts as a starting signal, for example, a well-defined leading edge. A restricted repertoire of characters therefore contains something of the order of 15 individual characters.

More sophisticated applications of character recognition techniques often demand the use of alphabetic characters as well as numerals which must include shapes specific to national (e.g. Dutch, Scandinavian) alphabets. Provision must also be made for 'accent signs', such as circumflex (^), acute (´), tilde (~). Extension of the range into lower case alphabetic characters and reasonable provision for the host of symbols used grammatically, scientifically, financially etc. leads to the conclusion that a repertoire cannot be regarded as comprehensive unless it contains over 100 individual characters.

A balance has to be struck between the demands to be met by the code and what is feasible in printing. Many office machines print ten characters per inch with vertical spacing of six lines per inch. In conventional letterpress work, the printers normally produce printed characters in a multitude of sizes, pitch and line spacing to meet the preferences of the customer and for practical purposes are largely unrestricted in regard to the size or spacing of characters.

The readability of characters is also a function of the consistency of the ink density and the degree to which the outline of the imprint conforms to specification. The difference between the dimensions of the printed character outline and those of the nominal imprint, depends on the printing technique. 'Halo' effects are characteristic of certain reprographic processes and every technique of impact printing results in an impression in which the contrasts are wider than the pattern on the type face. To a certain extent letterpress printing techniques also find difficulty in preventing 'squeeze' of the ink to the edges of the character outline. For office machines of the impact type e.g. typewriters, the more satisfactory method is the employment of ink ribbons with a homogeneous base. By this means an uniform amount of ink may be transferred in toto to the document being printed, without textural pattern as is derived from ink ribbons which have a textile base. Uniformity of ink density from one character to another is also dependent upon correct pressure at the type-face. In practice this is best achieved by commercially available encoders in the case of MICR and electric typewriters in the case of OCR. Nevertheless, all characters should be economically and technically practicable for use with all printers and the shapes of the characters should be as nearly as possible the same when printed by any method.

Placing of characters to meet machine needs

Character reading applications involve a very wide range of paper documents and printing processes. Document sizes may range from quite small cards for use as garment labels, for example, to pages of print or tally rolls. The engineering problems associated with moving the print on the document in correct registration past the reading station, more fully dealt with in Section 4, are simplified if there is only one such code line on a document: the earlier machines were and the cheaper machines are still single code line readers. Yet while many documents may carry only a single line of machine readable information others must carry several lines. Nevertheless there is one consideration in common: the area intended for reading by machine must be fixed in relation to certain reference edges or marks on the document.

These areas which are to bear the machine readable characters should remain at all times clear of other material which could interfere with the reading process. Many difficulties can be avoided if all the data is imprinted at the one time and therefore in one font but this may not be practical in particular instances. Confusion between the printed character and other marks in this 'clear band' may be minimized if magnetizable particles are included in the ink, or carbon black pigment inks are used for example. In these two cases respectively, reading devices sensitive to magnetic flux or to the high absorption of light from the near ultra violet into the infra red regions will discriminate the machine readable character from most extraneous matter in the clear band. Nevertheless most reading systems require the area around the characters to be read to be as clear as possible. Minimum line spacing must be sufficient to allow adequate separation between the highest character boundary in a line and the lowest character boundary in the line above, in addition to the height of the character itself. The width of the clear band specified is commonly as large as $\frac{5}{8}$ " (16 mm.).

Similarly, the spacing of characters along a code line needs to be uniform except where a gap is left for the purposes of dividing the line into fields of information. Usually it is desirable that the gap between fields should be a multiple of the character spacing also. Reading devices are usually very critical in this respect and may dictate tolerances as small as 0.015 inch.

Although reading devices are being developed with the ability to search a print area and reorientate a character pattern which is found to be misaligned vertically or angularly about the reference line for the character, this performance can still be achieved with only relatively small variations from the true reference line. The permissible character skew may be as little as $1\frac{1}{2}$ degrees and the vertical misalignment as little as 0.007 inch.

These demanding requirements are within the capacity of printers employing normal production methods but there is a great deal of justification for the claim of an American bank-cheque printer to the effect that the requirements of character recognition transformed printing from an art to a science.

Print quality considerations

Degradations of print quality on documents from whatever nominal

standards are set should be held to a minimum since this will have a bearing on the system performance. Either the printer or the reader may occasionally operate incorrectly and while the system can be framed to allow acceptable degrees of such failure it is usually necessary to determine the norm. Assuming effective maintenance of print and reader mechanisms it is probable that the printer is the more susceptible to preventative measures of control. Certainly pre-printed stationery should be subjected to inspection and control before entering a system; to a lesser extent encoders, typewriters and other office printing machines are susceptible to quality control of the print output within the system. Documents passing outside the system such as those handled by the public and returned, are not amenable to control of the final state of the print or paper and some method of inspection of the document on re-entry to the system may be desirable in order to weed out those which have been unduly affected.

In general the printed character should present as high a contrast as possible to the background of the document. It is recommended that for optical character recognition applications, inks with a high absorption from the near ultra violet to the near infra red regions be used: for magnetic ink character recognition systems the magnetizable material must be present in the ink in sufficient quantity and distributed sufficiently homogeneously for adequate and unambiguous magnetic patterns to be set up. In the letterpress printing process the ink, and consequently also the magnetizable constituent, is pushed to the edges of the print although this may not be visible to the naked eye. In severe cases the magnetic field will be stronger at the edges and cause secondary pulses which can make the pattern indistinct to the reader. For these reasons higher demands are made on the printing of magnetizable script than on that intended for optical character recognition.

Stroke widths and character outlines should be held as close as possible to the nominal values of the specification. It will be usual for tolerances to be specified within which the reading process will not be affected unduly by variation from the nominal values but these tolerances may be as low as 0.002 inch for MICR or 0.006 inch for OCR.

Within the stroke outline there should be no voids so great that they interfere with the reading process. Therefore some voids of a certain degree may be tolerable in particular systems, always providing that the character remains humanly readable. According to the application voids may be limited to almost as little as 0.015 inch in height for MICR and to 0.008 inch across for OCR.

Even more demanding perhaps of the conventional standards of printing is the requirement that there should be no extraneous ink spots within the clear band which can interfere with the reading process. The more sensitive the reader, the more demanding it is in this respect and it must be remembered that the same considerations apply to ink spots on the back of the clear band. MICR printing is likely to be prone to ink spots on the back of the clear band through 'setting off' from other documents. The maximum size of spots which are tolerable is usually between 0.003 and 0.008 inch in diameter according to the sensitivity of the reading mechanism.

Magnetic readers depend on intimate contact between the printed magnetic ink and the magnetic reading head. A printed impression which

depresses the surface below that of the surrounding paper by more than 0.001 inch is likely to cause the ink to be held away from the reading head to an extent which causes a deficiency in the signal wave form. Such depression of the paper surface is called embossment and as a general rule must not exceed that which a skilled person could barely detect by eye.

It must be borne in mind that however desirable background printing may be on the document, for example for security purposes, it is very likely to affect the contrast achieved between the background and the character. Optical character reading systems are more prone to be disturbed by this factor but background printing could also be a source of embossment troubles for a magnetic ink reading system. In both cases it is usually regarded as better omitted from the clear band or at least minimized by the use of only light coloured inks which are not so likely to mask extraneous ink spots.

Inspection of the printed characters by visual scrutiny, as is normal practice for printed output, is not sufficient for character reading purposes. The checking of measurements under a microscope may deal with dimensional considerations and where the machine readable documents are to be presented to the automatic reading device in virtually pristine condition, presentation of test samples to a standard reading device will indicate the acceptability of the product. More often the document has to survive considerable handling in use and the print specification for the new document must take account of this in order that the characters are still machine readable when ultimately presented to the reading device. Tests for resistance to abrasion on the part of the printed machine readable character may be desirable. For the printer to assess the quantity of magnetic material printed the most convenient method is measurement of the signal level, which usually has an allowable variation from as little as 25% to as much as 300% of the nominal value of a standard-height stroke. A comparable figure for OCR ink densities are typically, reflectance less than 60% of the bulk paper reflectance.

Paper considerations

The paper used in character reading applications should contribute to, rather than detract from, the distinguishability of the printed characters to the machine reading system without destroying its acceptability to the human reader or its capacity to withstand the handling conditions to be expected during its life cycle.

The paper should be dimensionally stable as far as possible and to this end printers have turned to air conditioning of the press shops in some situations. The atmospheric conditions at the time of printing should ideally correspond to those applying in the zone where the document will be machine read and commonly this means air conditioning to the standards required for computer operations.

It is a desirable feature that the surface of the paper shall be absorbent at least to an extent which will allow the ink to dry before setting off onto other documents.

Maximum contrast of the printed character with its background leads to a preference that for OCR applications the paper should be white, have a flat finish and a low gloss.

The inclusion in the body of the paper of material to which the reading device is sensitive is to be avoided as much as extraneous ink referred to previously. In OCR applications this takes the form of specified maximum dirt count, and for MICR applications there is usually total prohibition of magnetic particles in the paper.

The opacity of paper also enters into the degree of contrast obtained under the reading head in OCR applications. This is specified in relation to the background which exists in the vicinity of the reading head and which may therefore show through the document during the reading operation.

Fluorescent additives are sometimes used as brighteners for paper but should be avoided for OCR use as they may have a significant effect upon the reflectivity of the paper.

Water marks deserve special attention because they may offend both against the requirements for a flat surface and for minimum opacity and are better avoided at least in the clear band.

Mechanical properties of the paper may be demanded by the system outside the reading process. Tear resistance is relevant to general handling and may be an important consideration for the preparation of documents on line printers, in which circumstances the direction of the grain of the paper may be in conflict with the direction desirable from the standpoint of the paper transport of the reading device. The choice of the optimum combination of characteristics must be a matter for selection according to the particular application. Special mention should perhaps be made of the significance of fold resistance in papers used in OCR applications because of the tendency of folds to interfere with optical reading systems.

Progress to date in standardisation

Work has been proceeding for several years in the International Standards Organisation in an effort to produce standard MICR and, more recently, OCR fonts.

The position reached in early 1966 is that two MICR fonts have been accepted as proposed standards—one known as E. 13B (Fig. 16) which



Fig. 16

originated with the American Bankers' Association in 1958 and was sponsored by the American Standards Association, and the other known as C.M.C. 7 (Fig. 17) devised originally by the Compagnie Des Machines Bull of France and subsequently developed and sponsored by the European Computer Manufacturers' Association.

The E. 13B font contains numerals 0-9 and 4 special symbols, of nominal character height 0.117 inch and nominal character spacing 0.125 inch. The characters are composed of squares measuring 0.013 x 0.013 inch with rounded corners which give rise to the stylised appearance of the font. Its principal use is on bank cheques in the form of a single code line.

The C.M.C. 7 font comprises the numerals 0-9 the alphabetic characters A-Z in upper case only and five special symbols. There are four nominal character heights namely 2.7, 2.85, 3 and 3.2 mm with a common minimum inter character distance of 0.67 mm which leads to a variable character pitch according to the width of the characters concerned but may be as small as 3.17 mm pitch. The character is made up of 7 strokes enclosing 6 intervals of either long or short widths. It is the combination of these long and short intervals that determines the character code. The outline of the conventional character is obtained by the strokes being cut more or less to the shape of a conventional font. This font also is intended primarily for use in a single code line and the provision of four sizes of characters is aimed at the use of line printers, pica and other typewriters and letterpress type faces respectively from the smallest to the largest.

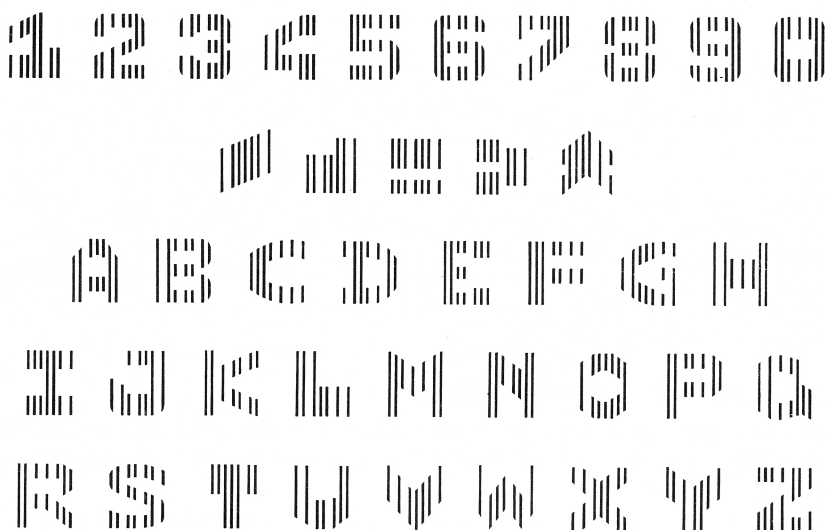


Fig. 17

Two optical character fonts, known as OCR 'A' (Fig. 18) and OCR 'B' (Fig. 19) are generally accepted in principle as proposed ISO standards. OCR 'A' is the more stylised of the two and was sponsored by the American Standards Association. OCR 'B' is a font which was designed to a commission from the European Computer Manufacturers' Association to produce characters as near as possible to conventional shapes while still retaining features necessary for machine reading.

OCR 'A' has been in use for some time in a numeric range of characters with some special symbols and for international standards purposes has been extended to include the 26 upper case letters of the alphabet plus 7 upper case national letters, 3 currency symbols, 16 assigned (recognised) symbols and 4 control symbols. This font is specified in 4 standard sizes, namely nominal heights 2.4, 2.84, 3.2 and 3.8 mm. Horizontal character spacing varies according to the inter character distance specified for each size and has yet to be sti-

1234567890
 JYH I ABCDEF
 GHIJ KLMNOP
 QRSTU VWXYZ
 ; , ' / { } % ? & \$

Fig. 18

A B C D E F G H a b c d e f g h
 I J K L M N O P i j k l m n o p
 Q R S T U V W X q r s t u v w x
 Y Z * + , - . / y z m ð ø æ
 0 1 2 3 4 5 6 7 £ \$: ; < % > ?
 8 9 [@ ! # & ,]
 (=) " ' ` ^ ~ ˇ
 Ä Ö Å Ñ Ü Æ Ø ↑ ≤ ≥ × ÷ ° ¤

Fig. 19

pulated. Line spacing is likely to be particular to each size of character but again has yet to be announced.

OCR 'B' font contains a total repertoire of 113 different characters, that is the numerals 0-9, 26 upper case letters A-Z plus 12 national letters, 27 lower case alphabetic characters, 2 currency symbols, and 36 assigned symbols. There are 27 lower case alphabetic characters because an alternative graphic 'm' has been given for use with those printing devices which have variable pitch. This font is specified in 3 sizes namely, nominal character height 2.4, 2.8 and 3.2 mm. A fourth size, 3.8 mm height, is likely to be added to the range for credit card purposes. Character spacing and vertical line spacing have not yet been announced.

In both the proposed standard OCR fonts the range of sizes is intended for use on line-printers, typewriters, other office machines and credit cards from the smallest to the largest respectively.

Progress envisaged

The most immediate and hopeful prospect is for the design of fonts to be standardised internationally and there is every reason to expect that this will be achieved on the lines of the previous passage in the not too distant future.

Given the opportunity to develop upon known standards for the fonts it is hoped that character reading devices will become more efficient and ultimately cheaper.

These may at some stage be supplemented if not supplanted by machines capable of reading a mixed selection of fonts.

Ultimately machines will surely become available with the ability to read handwriting even though the script may have to be constrained to within the designed capability of the machine reader. This should open up whole new fields of data capture at its place of origin, including the private individual in his home.

SECTION 4

Document Handling and Transportation

DOCUMENT HANDLING AND TRANSPORTATION

Introduction

The reader will already have perceived from Section 2 of this book that paper is a somewhat unstable medium which is subject to variations of dimensions and rigidity due to changes in relative humidity, and is easily damaged by careless handling. Each particular character reading application has its own paper handling problem which may range from the closely controlled in-house situation, in which the documents are never handled by personnel outside the system, to the reading of turn-around documents which are used by the public at large. This latter situation obviously presents the greatest problems to the designer of the equipment, and it is likely to have the poorest performance however ingenious he may be. The engineering techniques employed by the various manufacturers have been extensively patented and this is to some extent responsible for the differences between machines which are designed to do similar jobs.

The earliest reading machines required the documents to be fed manually one by one into the input section. The operators withheld unsuitable documents from the machine and hence reduced the risk of reading errors and paper wrecks. In the case of the earlier machines which could only read one line of information on a document this was clearly an extravagant use of operating staff, but for the page reading machines which are now appearing there may be applications where the data available from each document is sufficiently great for the hand feeding system to be economic.

For the main part, this section of the book considers the automatically fed document transport, since it is in such machines that the document handling problems are most severe. Machines which handle continuous stationery, such as tally rolls, are simpler and are given separate consideration at the end of the section.

The basic requirements of a paper transport

These may be simply defined as follows.

- 1 To be easy to load with documents.
- 2 To pick off one document at a time from the input pack.
- 3 To present these individually picked documents to the reading station in a fully controlled manner.
- 4 After reading to stack the documents for easy unloading without upsetting the 'facing'.
- 5 To handle the documents without damaging them.
- 6 To be as tolerant as possible of mutilated documents, staples and pins.
- 7 The machine must be safe and simple to operate and jams must be easy to clear.
- 8 The machine should be easy to maintain.

Outline of an automatic document handling system

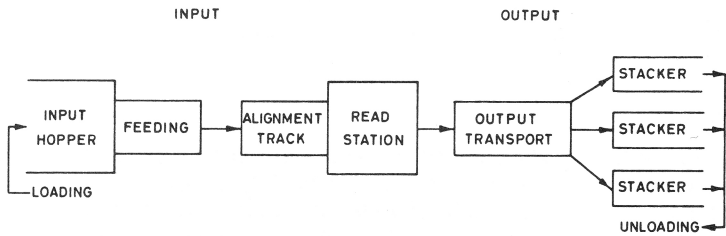


Fig. 20

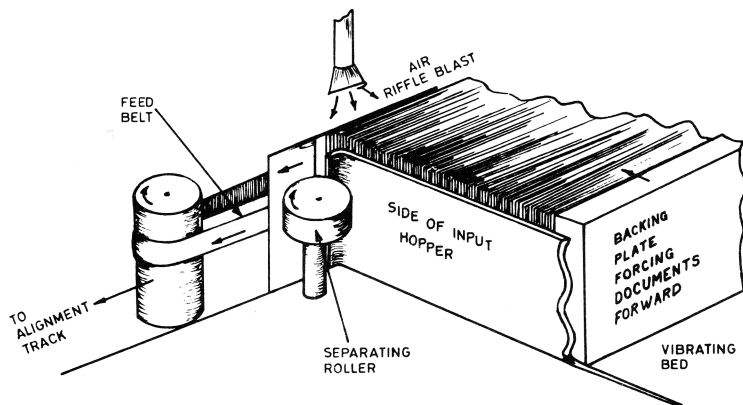
Fig. 20 is provided for the benefit of the reader who is unfamiliar with the basic elements of a document transport system. This illustrates in block diagram form the principal parts which are discussed in this section of the text.

Loading. All machines so far introduced require that the documents are 'faced' the same way, that is to say all the documents in the input pack must be correctly orientated with the surface to be read facing the same way. The facing of the documents is usually carried out by a supporting team of sorting staff who also remove staples and pins, ease out folds and creases in forms, and remove those forms which are badly damaged.

The machine operator will then place a suitable pack in the input hopper of the machine. The input hoppers take various forms, but in general they comprise a box-like compartment which is open on one side for loading. This hopper incorporates some means of pressing the documents towards the feeding mechanism. Position, orientation and capacity of the input hopper are factors of interest to the user. In some cases the documents stack on edge, in others one upon the other. In considering the convenience of any particular input hopper the document size is a significant factor for what can be quite satisfactory for cheque size documents may be unmanageable for quarto forms.

In any automatic feed system, it is desirable to be able to replenish the input hopper without stopping the machine. When this is being done it is important that the conditions prevailing at the feed end of the input pack are not disturbed, and this generally means that the pressure with which the documents are pressed against the feed end is not changed. If this pressure is due to a direct gravitational effect it is usually only necessary to put more documents on top of those already in the hopper. However, many systems employ a spring loaded, or servo-motor loaded feed, and in these cases the hopper is usually equipped with two hinged feeder plates, either of which can force the pack onto the feed point. To replenish the hopper the operator introduces the next pack of documents behind the plate which is pushing the existing pack. The spare feeder plate is then drawn back behind the new load and set to push it forward. The other feed plate is then withdrawn, permitting the documents to join up to form a combined pack.

Feeding. At the feed end of the input hopper the documents must be separated and fed one at a time. Mechanization of the typical means which human beings employ is the general line of approach. Air blasts may be used to riffle them apart, rubber belts moving at different speeds rub them apart, and a curved document path is sometimes used to assist feeding off the pack. See Fig.21. The separating roller shown here moves slowly in the opposite direction to the main feed belt thereby pushing back and parting double fed documents.



ILLUSTRATING A DOCUMENT FEEDING ARRANGEMENT

Fig.21

The separated document is often lifted off the feed pack by a vacuum technique in which a perforated belt, drum or feed shoe grips the document. Such an arrangement can be adversely affected by perforations in the documents, which line up with the vacuum ports; this condition is obviously avoided by careful document design when such a mechanism is chosen. The friction belt feeds, which are an alternative to vacuum systems, do not have this shortcoming. The riffling air blasts which are used at the feed point are usually a source of noise.

When connected to a computer for on-line working, it is usual for the computer to control the feed. This gives rise to the so-called 'on-demand' mode of working in which the document is fed only when the computer is able to accept the reading arising from it. The document feed is initiated by switching on the feed vacuum or starting the feed belts.

The document transport has to present the characters to the read station within the prescribed positional tolerances. The printing (positioned relative to the reference edge of the document), will already have been carried out with this in mind. The document transport will therefore aim at aligning this reference edge of the document with the reference edge of the reading station. In some machines this lining up is accomplished in the feed hopper and feed station by using a vibrating base in the feed hopper which joggles the documents so that the reference edges rests on the hopper base plate.

At the feed station the document pick off and feed arrangement is carefully designed to prevent any skew occurring as the document leaves

the feed pack. However, it is more common for the document to be aligned in the transporting part of the machine; this is described later.

The condition of the documents can have dramatic effects upon the performance of the feed mechanism. Items in mint condition can be extremely difficult to separate, but once separated they can feed well. Perforations, severe folds, creases, staples and pins can impede feeding by causing documents to catch on one another. Damaged leading edges and reference edges are also common causes of difficulty. Shortcomings in the document condition are usually less noticeable if the grain runs in the direction of travel of the document in the machine, and since the documents usually travel in the direction of their longest dimension such a grain arrangement is known as long grain. However, system requirements (e.g. printing of turn around documents on continuous stationery by line printers) sometimes make it difficult for the user to produce documents with the long grain in the desired direction (See comment below on output stocking)

The feed unit itself will cause wear on the documents, becoming catastrophic if the feeder jams and the documents become wrecked. Here a sensitive jam detector is invaluable and various techniques have been developed. Probably the most effective is that which stops the feed within a feed cycle, and one such system is based on checking that when a document is called for it does in fact leave the hopper within a feed cycle period. Any device which is slower than this will give rise to more extensive document damage.

Apart from catastrophic wear all feed systems impart characteristic wear on the documents. Forced bowing of the document to aid separation and frictional effects of belts and drums all leave their evidence. Sometimes a friction mark will be quite significant and on optical character recognition systems this may be troublesome. Fortunately, the positions of such belt marks are well defined and it may be possible for the user to avoid these areas when choosing the position for the characters to be read.

In the feed area it is possible for the moving parts to have sufficient force behind them to cause injury if operators fingers are inserted. Manufacturers being well aware of this, take steps to minimise the risk. Clutches are sometimes fitted to limit the power available and emergency stop buttons are usually situated convenient to the feed point.

Transporting the document to the reading station. The length of this part of the machine may be a few inches or a few feet. In it the document is conveyed on roller or belt tracks at speed as high as 22 ft. per second (15 mph). At these speeds it is essential that the leading edge of the document should at all times be guided and this is achieved by driving it between pairs of rollers, pairs of belts or rollers and balls. It is also desirable that the document experiences no retardation in its motion and in some designs a degree of acceleration is built in to ensure that this does not occur.

As the document traverses this part of the machine it must be prepared for alignment against the reference edge of the reading station, unless this has been done earlier. It is in the latter case that this part of the machine is short. In the former case alignment is induced by

creating a transverse force on the document to urge it against a reference edge on the conveyor. See Fig.22

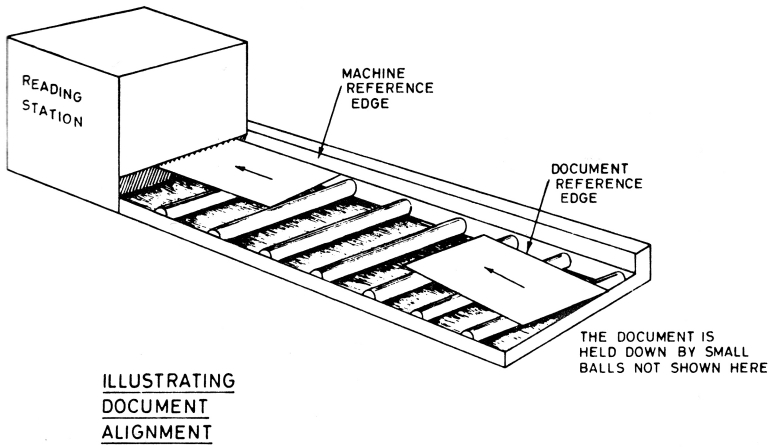


Fig.22

The transverse force is typically produced by skewing the conveyor rollers. The paper is held onto these rollers by spring (or gravity) loaded balls which permit the required degree of sliding as the document moves into alignment. By the time the alignment process is completed the document should be travelling at a constant velocity as it is launched into the reading station.

The alignment forces are usually quite significant; they have to be. This may result in forms which are nearly square in shape being rotated or skewed in the alignment process and it is therefore, usual to specify a limiting length to width ratio for the document. The risk of failure due to skewing (skew failure) increases as this limit is approached.

Skew failure is the condition in which the document passes into the reading station with its reference edge not parallel to that of the reading station. The reading station will usually tolerate some skew which will be shared between the printing skew and the transport skew. From time to time both may be additive and give rise to reading failures which at first glance are difficult to understand. Only careful examination of the evidence will show where the trouble lies and all too often the evidence is of a transient nature.

Associated with this part of the transport one usually finds a double document detector. This may employ optical, differential vacuum or thickness sensing techniques. In some optical systems two or more photo-sensing stations along the document track check the document length and by means of a logic circuit determine if it is too long, and hence deduce that an overlapping document is present. Other optical systems make use of the difference in opacity of one and two thicknesses of document. The differential vacuum detector presents a vacuum port to each side of the document. If there are two documents present they can part and seal both ports and this is readily distin-

guishable from the single document condition. Thickness sensing employs a microswitch which is set to detect anything over one thickness of document. Such an arrangement can be tedious to set up and may be liable to false alarms.

Microswitches are, however, very effective for the detection of jams. Photo-cell detection of jams is also employed on some machines, the jam being detected by logic which checks the transit of the document from photo-cell to photo-cell along the conveyor path. It is important that tractive power be removed from the documents as soon as a jam is detected. The faster the speed of the transport the more difficult it becomes to do this quickly enough. The most successful techniques employ a mechanical means of opening the conveyor to take the drive forces off the paper.

The performance of the document transport is considerably affected by the document condition. Pins, staples, creases, folds and crumpling can all give rise to unmanageable documents. Folded over or damaged leading edges are likely to cause wrecks and damaged reference edges are likely to give rise to skew conditions.

Apart from severe damage to documents involved in a jam the action of the belts or rollers of the transport usually confers negligible wear pattern on the documents. As very little acceleration occurs over this part of the machine, only small forces are applied to the document so the wear pattern is not usually significant.

Operator safety is again important since although the transport is normally closed when it is in operation it must be opened up for the clearing of jams. The necessary safety is usually achieved by interlocking the motor drive circuits with contacts on the covers of the machine.

Output stacking. Having passed the reading station the document is conveyed to the output stacker on a roller or belt conveyor as previously described. There may be one or several output stackers, the latter arrangement being commonly used in the MICR cheque handling machines. For computer input of OCR one can expect to see machines with two or three output stackers assuming there is no need for physically sorting the documents.

The documents are usually diverted into the required output channel by a diverter blade which is interposed in a gap between the sections of belting or rollers and diverts the document to one side into a branch conveyor. This conveys it to the output stacker.

At the stacker the documents must be brought to rest and stacked neatly at rates of up to 25 a second. This presents an aerodynamic problem as the leading edge has the minimum of constraint upon it here. At this speed there is no time to be wasted. The documents must be slowed down as gently as possible and various arrangements have been developed to minimize damage to the leading edge. Having stopped the document the tail end of it must be quickly moved out of the path of the following document. Here again, various systems, pneumatic and mechanical, have been devised.

The quality of the paper used has great effect on the output stacking and it is here that the benefits of long grain paper are most noticed. The extra longitudinal strength which long grain gives to the document

is of special value in withstanding the deceleration forces. Some machines seek to give the document increased longitudinal rigidity by inducing a lengthways bow to it as it enters the output stacker. Output stacking also introduces its own wear pattern characteristics on the document. These are principally damage to the leading edge and bowing along the document length, and the extent to which these are reduced is a direct indication of the success of the design in tackling the deceleration problem.

The transporting section leading to the output stacker requires jam detection and safety interlocks on access covers just as on the earlier parts of the machine. At the output stacker itself the drives should be harmless to operators fingers for it is usual to unload these stackers without stopping the machine. Sometimes an unloading interposer plate is provided to part off a portion of the pack for unloading.

In all cases provision must be made to stop feeding into the output stacker when it is full. This can be achieved by stopping the machine, or switching the output to a spare stacker. Eventually however, operator intervention becomes necessary and it is helpful if warning of a full stacker condition can be given before the machine is stopped on this account.

The user will be particularly interested in the position, orientation and capacity of the output stackers since these features can considerably influence the operating efficiency.

Speeds of operation

The higher speeds of operation are at present mostly found in cheque handling machines using MICR. Six inch long documents are handled at a rate of up to 1600 a minute involving a conveyor speed of about 22 ft. per second. To achieve this the paper must comply with an exacting specification and the documents must be in reasonably good condition.

The actual throughput speeds achieved on document transports are not always self-evident. The simplest case is that in which the feeding is performed cyclicly. In such a machine one document is fed in each cycle and feed rate is not affected by the document length. However, most machines will handle a range of document sizes and to restrict oneself to a cyclic feed system results in a long gap between short documents. This has given rise to acyclic feeding in which the feeding of a document is timed from the trailing edge of the preceding document. Manufacturers usually quote a simple formula from which users can derive prospective throughput rates for the document sizes they have in mind.

At low speeds of operation the document transport becomes less demanding on document condition. Hence for the situation in which a user anticipates considerable wear and tear on the document before it reaches the document reader a slow speed mechanism is likely to give better results.

The emphasis on high document speed is not so great in page reading machines. Frequently the document could be moved past the reading station more quickly than it could be handled by the reading system. Hence with page reading machines the paper transport problem is less

severe. Typical page rates at present are 1 to 5 pages a second, depending on the number of lines to be read.

The large size of the document in page readers tends to present problems of its own. The mass to be accelerated and arrested is greater and the wider leading edge requires greater constraint when in motion, so that it is perhaps fortunate that the document velocity does not need to be so high. As reading speeds advance it remains to be seen to what extent paper transport designers can master these problems.

Static electricity

No consideration of mechanised paper handling is complete without some comment on static electricity, though in a well designed machine operating in the correct environment it should be troublesome.

Static electricity is generated when a material of poor electrical conductivity is rubbed or rolled with another material which may be conducting or non conducting. Hence, paper travelling over rubber belts or metal rollers is prone to develop static electricity charges. The faster the motion the more pronounced is the effect.

When the paper becomes charged with electricity it tends to stick to some part or other of the machine, or to other sheets of paper. This, of course, can seriously upset the paper motion. The phenomenon is likely to appear first at the output stacker where the document, released from the constraining influence of the conveyor, fails to stack and causes a jam or, maybe, is propelled into the room.

If the documents are stabilized to an environment having a relative humidity of 50% or more the moisture content of the paper is high enough to cause the static electricity to leak away as quickly as it is formed. Exceptionally however, local heating of the paper under optical reading stations can have a drying effect permitting static electricity effects to manifest themselves.

Various brush like devices which rub across the paper to conduct away electrical charges have been developed but the choice of the wrong material can introduce more static charge than it removes. The best course is the avoidance of very dry documents and very low relative humidity and attention to these points is probably the most effective line of control.

Continuous stationery readers

The document in this case is in the form of a roll of paper, typically a tally roll. The document transport is tailor-made for the size of roll and paper width to be used and the paper handling problem is much easier than for the individual document machine. The paper is usually respooled as it leaves the reading station, and being in continuous form it can be stopped and reversed for re-reading quite easily under program control from the computer.

The paper speeds are usually slow, of the order of 6 inches to 12 inches per second. Registration against the reading station reference edge is dependent upon alignment of the paper from feed spool to take up spool. Creases or folds in the roll can upset this.

Both the type of paper movement and the speed make it unlikely that

this type of reader will experience trouble from static electricity, though if it does it will probably appear on the output spool.

The paper used in these machines is usually poor quality and consequently quite frail and care is necessary in handling it to avoid creases or edge damage. Typical systems employ printing mechanisms attached to cash register machines and the spools can be unloaded by suitably instructed staff. It is helpful to have containers available for protection of the spools in transit.

Maintenance

In choosing a machine the prospective user should carefully consider its maintenance requirement. These machines invariably become dusty because fragments of paper lint accumulate inside the machine as the documents are conveyed. The amount of this dust is greatest for the fast machines and is also modified by the quality of the paper and condition of the documents. The first victims of dust are the photo-electric devices and these may need cleaning quite frequently, maybe every 4 or 5 hours of running. This may be an engineering or an operating routine depending upon the machine manufacturer.

General cleaning to remove dust from the inside of the machine and the cleaning of filters in the pneumatic and vacuum systems is also necessary quite frequently and will be done by the maintenance engineer on a preventive routine basis. In the high speed machines moving parts require preventive maintenance inspection and replacement at intervals specified by the manufacturers. These are not usually very frequent and can be planned to cause no inconvenience to the management. Probably the most common mechanical failure causing inconvenience is a broken conveyor belt, and this is usually quickly put right, providing that a spare belt is to hand.

The page and continuous stationery readers are least demanding in their maintenance because their document transport systems are simpler and the paper movement is slower. Consequently the mechanical parts are less likely to fail and the dust problem is much reduced, resulting in an easier maintenance situation.

SECTION 5

Techniques of Document Scanning and Character Recognition

TECHNIQUES OF DOCUMENT SCANNING AND CHARACTER RECOGNITION

Introduction

It is the purpose of this section to catalogue briefly the various techniques used in character recognition equipment without direct reference to particular manufacturers devices. Both currently exploited techniques and those promising techniques currently under development and on the horizon will be covered. Approaches which have proved ineffective will not be described. For fuller and more formal descriptions the reader is referred to Stevens (1961) which contains an extensive bibliography and also comments at length on many technological and systems aspects of CR. Also Bijleveld (1963) reviews a number of systems with particular reference to those using stylised character fonts.

Black-box model of a pattern recognition machine

In order to establish a terminology for the description of techniques used in CR devices it is useful to break down a general model of such a device into its component 'black boxes' Fig.23 and then to describe techniques which can be used in each box. It should not be assumed that these boxes are independent of each other in any real sense.

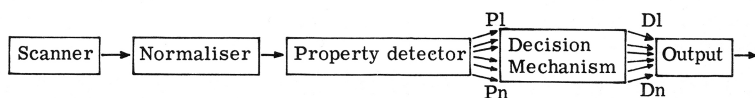


Fig. 23

The scanner is the transducer, a magnetic or photo-electric device of some sort, which converts the presented character into suitable electrical signal(s) for processing by electronic apparatus. The normaliser, if present, modifies the signal from the scanner to obtain a processed version of the input signal more suitable for detailed analysis. Such processes can include operations for standardising the amplitude and position of the character signal, improving the contrast and quality of the character and transforming it into a two-state (binary) signal. Normalising is essentially a 'quality improving' operation and should not alter the fundamental shape of the character. Ideally the normaliser would convert a distorted input pattern into its intended 'perfect' form. The degree of which this is achieved offers a basis for comparing techniques.

The property detector processes the standardised signal to extract from it a set of characteristic properties on the basis of which the character can be subsequently identified. Ideally the property detector response would be invariant under conditions of varying quality, style, position, etc. of the presented character—thus the tolerance of given property detection techniques to such variables is a useful basis for comparison.

The decision mechanism, on the basis of information provided by the property detecting device(s) alone, has to decide upon and indicate the most probable identity of each presented character. If it is not possible for the decision mechanism to arrive at an un-ambiguous decision then that character is indicated as being unreadable, i.e. rejected. Again the tolerance of a decision mechanism to variation in the property statements fed to it form a basis for comparison of the various decision techniques.

The response of the decision mechanism is fed to a suitable output unit which records the responses in an appropriate form.

The heart of a CR system is clearly the three central units, normaliser, property detector and decision mechanism; the scanner and output units are essentially performing conversion operations only and do not in any sense process the signals flowing through them. The various components of a CR system outlined above will be discussed more fully below.

Classes of character recognition systems

Again for ease of description it is convenient to divide the general field of CR devices into the three categories indicated earlier in this handbook.

- (a) mark reading
- (b) magnetic ink character recognition (MICR) and
- (c) optical character recognition (OCR).

The different techniques used in these three categories will now be discussed in terms of the 'black box' model.

Mark reading

These systems use forms pre-printed to the extent that 'boxes' or 'windows' on the form have coded meanings. Information is recorded by making a distinct mark, either manually or by printer, in these windows. In the event that a mark is made in error it may be deleted either by rubbing out or by obliterating. In the former case the reading device detects that the signal is both low in level and gradual; in the latter case the reading device detects that there is not one mark but effectively a pair of marks and interprets this as having the meaning of no mark.

Since mark reading systems generally require several windows to allow a range of quantities (numbers) to be represented the reading device has to scan a large area of paper on which there may well be extraneous pencil marks, thumb marks etc. Most machines therefore detect three states—no mark, doubtful mark and mark. The detection of a doubtful mark may be used to reject the document being read.

The scanner for these devices is usually a row of photo electric cells measuring the artificially provided light reflected from the document. Each cell is usually assigned to one column of windows and detects the presence of a mark when the reflectance falls. If the reflectance falls quickly enough and far enough then a mark has been detected. If the change in reflectance is slow or too small then the mark is considered doubtful.

Although the document reject rate tends to be rather high mark reading has found many applications in data capture such as census returns, meter reading and order forms.

There is another form of mark detection called mark sensing where the conductivity of graphite pencil marks is utilised. However, this requires the use of soft pencils and is not applicable to printed outputs or pens. The reject rate is generally higher than that for mark reading.

Magnetic ink character recognition

All magnetic ink systems use highly stylised character shapes (c.f. E13B and CMC7) printed in ink loaded with magnetisable particles and are recognisable by machine from their vertical projection alone. Vertical projection reduces the two dimensional pattern seen visually to a one dimensional pattern (voltage waveform) for machine recognition. This reduction in the dimensionality of the pattern with consequent loss of information about detailed vertical structure accounts for the peculiar styles and high quality of printing required for use in MICR systems. Although designed to be recognised primarily from their vertical projection at least one system has been developed for recognising E13B using the complete two dimensional information available.

Scanner

The majority of MICR devices (using either E13B or CMC7) employ a slit scan to scan the code line containing the magnetised characters as they pass the sensing device at uniform speed. This sensing device is basically similar to a conventional tape recorder head which converts any variation in magnetic flux (due to the presence of magnetised material) across the head gap into a related electrical signal. Using a magnetic head with a vertically oriented gap long enough to cover the printed code band the output from such a head is proportional to the amount of magnetic material passing horizontally before and in contact with it, and hence to the vertical projection of characters printed in the code line. Non-magnetic printing gives no output and over stamping or printing leaves the electrical signal unchanged.

Property Detection

Due to the different structural nature of E13B and CMC7 characters the methods of analysis and recognition of the two styles are fundamentally different. The ink distribution in E13B characters is continuous in both horizontal and vertical directions, the vertical projection of the character is therefore also continuous and these characters are recognised from the varying analogue waveform which they induce in the scanning head. CMC7 characters on the other hand are constructed from discrete vertical stripes of ink with either of two permitted spacings between them (Fig. 24). Only the printed stripes contain magnetic material and the output induced in the scanning head is essentially digital in nature—each stripe producing an output pulse, recognition is based upon the spacings of the pulses so produced.

Property detection in the case of E13B becomes the measurement of the match or correlation between the shape of an incoming signal and

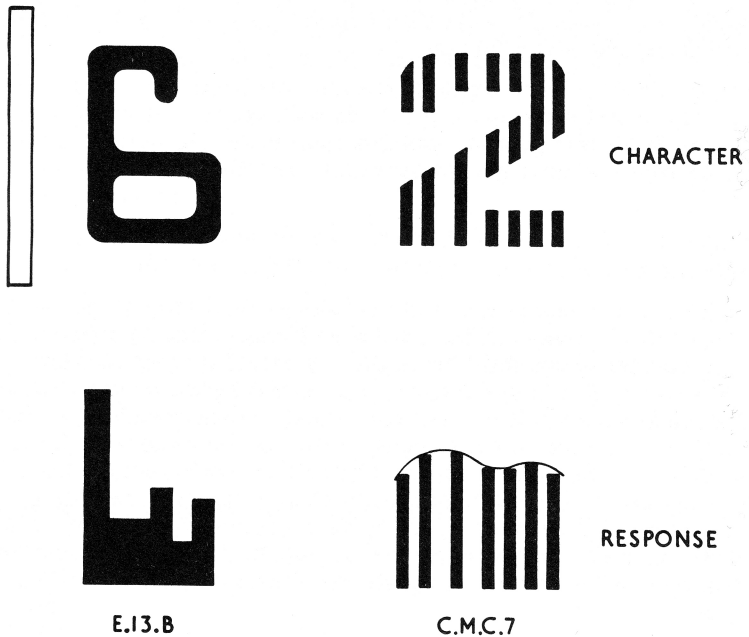


Fig. 24 Slit Scan

a set of standard waveforms. This comparison operation is usually performed by the 'matched filter' technique.

Systems for reading E13B as two dimensional characters are essentially the same as optical systems employing a divided slit scan and peep-hole mask technique.

For CMC7 property detection amounts to no more than converting the incoming train of variably spaced pulses into a conventional digit code indicating, as a string of ones and zeros, the sequence of long and short pulse intervals. The CMC7 numeric characters have been constructed so that the codes representing the printed characters differ from each other by at least two elements (digits) to reduce the possibility of error. A character producing an inadmissible code is automatically rejected.

Decision Mechanism

The decision mechanism employed with E13B characters simply detects the largest correlation or match between the incoming signal waveform and the representations of standard character waveforms within the machine. To avoid possible erroneous response signals arising from 'cross talk' between adjacent characters the maximum correlation response is only sought when it is known that an individual character is in approximately the correct position for recognition. This condition may be detected by observing that the gap between characters contains no magnetic material so that the vertical projection of the character falls to zero. This condition is detected in the output from the scanning head and the maximum response occurring within a short

interval following this event is taken to indicate the identity of the character just scanned.

The groups of pulses produced by CMC7 characters are digitally matched against a set of standard groups representing each possible character class. Characters are identified only if an exact match occurs, if this does not happen the character is automatically rejected.

Comment on MICR systems

MICR systems employ character styles designed expressly for machine recognition with only minor concessions to human legibility. In consequence MICR readers are comparatively simple devices and therefore, reasonably cheap. They are, however, very dependent on a high quality of printing being maintained if low reject rates are to be obtained. Although intended for reading magnetically there is no theoretical reason why they should not alternatively be read optically. Optical reading would of course destroy the immunity obtained in magnetic reading to over printing and surface disfigurement. MICR is now fully developed and, although marginal improvements in equipment will continue for some time, no major change in performance is likely.

Optical character recognition

OCR systems employ more or less conventional character styles printed in conventional, usually black, ink. Information about their presence and structure is obtained on the basis of measurement of optical reflectance alone, the source of illumination varying from system to system, some employing selected parts of the spectrum to enhance the contrast between character and background and to overcome by varying degree the effects of over printing and disfigurement.

The various 'black-box' components of OCR systems which have appeared in commercial equipment are described in limited detail below. A later section will deal briefly with techniques currently under development and likely to be incorporated in future systems.

Scanning methods

The purpose of the scanner is to convert the two-dimensional characters into an electrical form suitable for the property detection processes with minimum distortion. The scanning method employed depends both on the requirements of the property detection system and the paper handling mechanism employed. In many scanning systems the paper motion itself forms an integral part of the scan.

Slit scan Fig.24

As explained in connection with MICR readers suitably designed character styles can be recognised from their vertical projection obtained by scanning them horizontally with a long (tall) narrow slit. A measure of the aggregate reflectance (or magnetic material) within the slit effectively describes such a projection. In practice the characters may be either illuminated by a narrow strip of light and the amount of light reflected measured with a photo-cell or an image of the character may be projected onto a slit mask, a photo-cell then measuring the amount of light passing through the slit. In either case a constant horizontal relative motion between the character and the slit is introduced

to generate the complete vertical projection of a character. The resulting waveform is then delivered to the property detection and decision units with or without intermediate normalising. Vertical registration of the characters need not be closely controlled since the scanning slit can be made considerably taller than the individual character heights.

This is the simplest form of scanning, the relative motion between character and scanning slit is usually obtained by transporting documents to be read horizontally before the slit at constant velocity. Since all detailed information about the vertical structure of the characters scanned is lost in forming their vertical projection then the range of characters which can be recognised is limited and the tolerance to print degradation is small. Systems employing slit scan are therefore unlikely to find application other than in reading single code lines containing a limited range of characters (numerals plus a few control symbols) from individual documents. Speed of slit scanning is limited by the document transport which currently gives typical effective reading rates of the order of 1000 ch./sec.

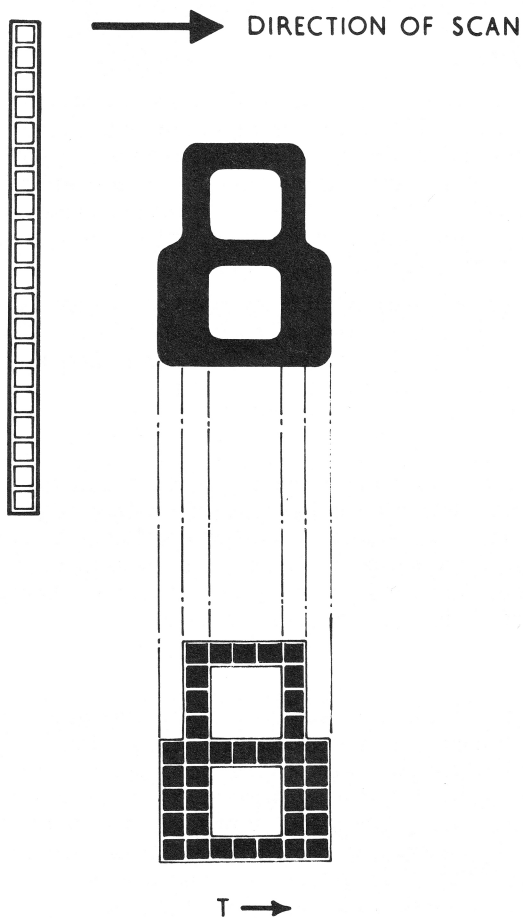


Fig. 25 The divided slit scan.

Divided slit scan Fig. 25

At the cost of losing the tolerance to vertical registration a vertical column of photo-electric devices can be used as a divided slit scan to obtain information about the detailed vertical as well as horizontal structure of characters scanned. Document motion may be used to provide the horizontal scan component as the character bearing strip of the document is imaged onto the column of photo-cells. (It is possible to obtain information about the vertical structure of magnetic characters in a similar way if a magnetic head made up of a stack of individual narrow track heads is used, as in multitrack tape recorders.) The number of individual elements used in a divided slit scan depends upon the resolution required in the vertical direction but is unlikely to be less than say seven cells in a character height. In order to work satisfactorily in a low print quality situation the vertical resolution may need to exceed twenty elements in a character height. If the divided slit method of scanning is to accommodate vertical misalignment then the column of cells must extend over the range to be encountered. Changes in the response from the scan due to vertical misalignment have to be corrected in the normalising or property detection unit. Systems using paper motion as an integral part of the scan should realise speeds quoted for slit scan.

Mechanical scanner Fig. 26

In the mechanical scanner the area to be scanned is brightly illuminated by some convenient light source. The illuminated area is projected by a lens system onto a rotating disc, around the periphery of which is a pattern of small uniformly spaced apertures. As the disc rotates the apertures traverse the area vertically, successive apertures are displaced radially by an amount equal to or slightly greater than their diameter so that in the course of a complete revolution a two dimensional area is exhaustively scanned. (In some equipments the horizontal scan component is given by horizontal document motion. The apertures in the disc are then arranged uniformly around a circle.) The reflectance of the character at all points in the scan is measured by a photocell which collects the light passing through the disc apertures. One rotation of the disc is sufficient to completely scan a character which is then represented by the signal waveform from the photomultiplier. This waveform is conventionally converted to a two state (binary signal and stored in a convenient storage device, e.g. two dimensional shift register. The dichotomising threshold applied to the raw scanner output can either be preset manually or can be adjusted automatically according to information gained from a preliminary scan of the character. Exploration of a document format to locate specific characters is possible using an auxiliary system of oscillating mirrors to move the field of view of the scanner to desired locations. Fig. 26 shows a Nipkow disc scanner with circular apertures as used in early t.v. systems, an alternative method forming square scanning apertures intersects radial slits round the disc periphery with a fixed slit at right angles. This method has been more extensively used in OCR to date than the Nipkow disc. Though relatively simple and inexpensive mechanical scanners are limited in their scan rate with a practical upper limit of approximately 400 characters scanned per second. If a high definition scan is required the disc may become uncomfortably large and difficult to stabilise. If high speed oscillatory

and irregular motions of auxiliary mirror systems are included these will require to be well maintained.

This system of scanning has to date been more widely exploited than any other in OCR.

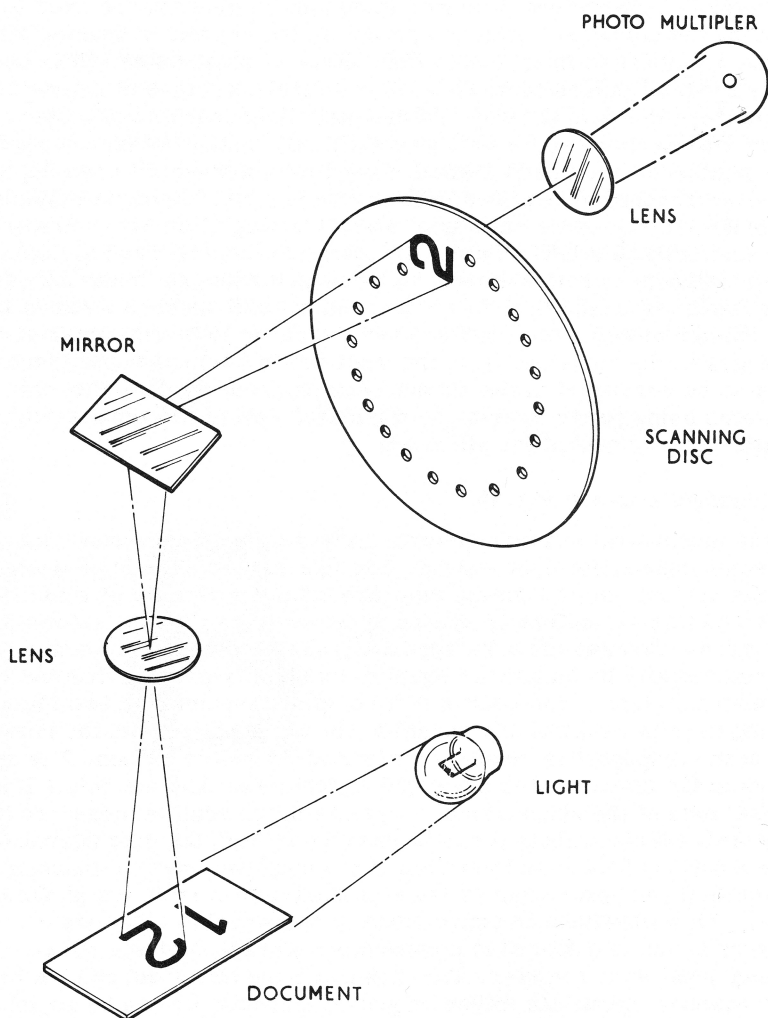


Fig. 26 One method of mechanical scanning (Nipkow disc).

Flying-spot scanner Fig. 27

The flying-spot scanner is basically similar to the mechanical scanner described above except that the position of the light source used to illuminate the document and the photocell measuring the light admitted by the scanning disc are inter-changed. Thus only a small area of the character is illuminated through the holes in the scanning disc at any instance. In the course of one revolution the whole character is illuminated point by point in sequence. The photo-cell is placed so that it collects and measures the amount of light reflected from the charac-

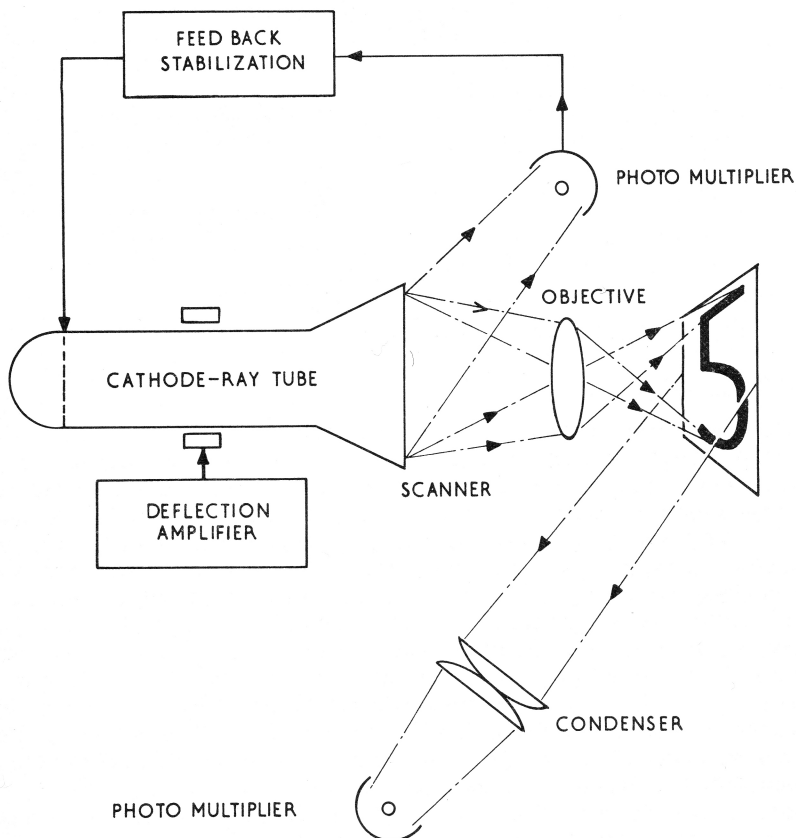


Fig.27 The flying spot scanner.

ter. The signal from the photo-cell is closely similar to that obtained in the mechanical scanner and there is no advantage in this arrangement while a mechanical system is used, in fact there is a disadvantage, since the character is illuminated from a scanning light source all ambient illumination at the character must be eliminated in order to obtain satisfactory performance.

In practical flying spot scanners the rotating disc is replaced by a cathode ray tube on the face of which a rectangular raster scan is excited which is used to illuminate the character—a photo-cell collecting and measuring the reflected light as before.

The scanning rate obtainable with a flying spot scanner is limited only by the time required for the light emitted from the c.r.t. phosphor to decay after the scanning spot has moved on. This time can be as little as $\frac{1}{10}$ micro seconds. Scan rates up to several thousand characters per second have been attained successfully. In addition, using a precision high resolution scanning tube, character selection anywhere within a 6" × 6" area can be achieved by electrical deflection of the scan raster without mechanical movement of either optical components or the document.

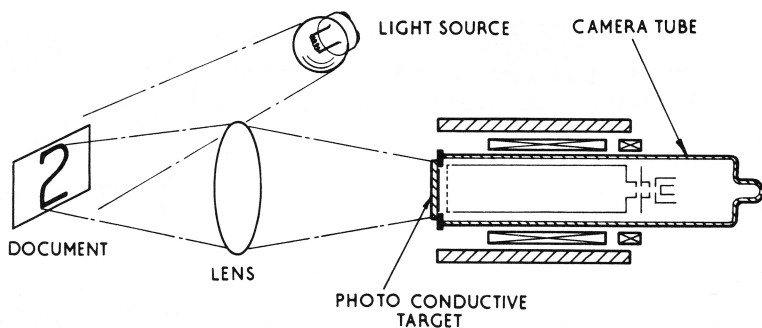


Fig.28 A t.v. camera scanner

T.V. camera Fig. 28

With the advent of small vidicon t.v. camera tubes which are relatively cheap, stable and durable these have been employed as scanners in some character recognition systems. The characters to be read are imaged onto the sensitive photo-conductive target of the vidicon camera tube. The resulting electrical charge pattern developed on the inner surface of the target is explored by a scanning beam which traces out a rectangular raster. The resulting electrical signal representing the character is closely similar to the output of a flying spot scanner and may be processed in the same way. Camera tube scanners have some novel features not present in any other types. The projection of a pattern of light onto the photo-conductive target sets up an equivalent pattern of electrical charge on its inner surface. This charge pattern persists, even if the incident light pattern is removed, until the scan beam, discharges it. It is therefore possible to store a pattern for an unspecified interval of time to be scanned out subsequent to the exposure of the target. Using a high intensity light flash of short duration to illuminate the document area to be read it is possible to read moving documents. The short flash effectively freezes the paper motion the resultant brief exposure is then stored on the photo-conductive target for eventual scanning. The act of scanning usually destroys (erases) the stored charge pattern and prescanning is not possible, however techniques have been developed using modified tube operating conditions (reduced scan beam current and modified camera amplifier) which permit repeated scanning of the stored charge pattern with only gradual erasure. A disadvantage of currently available camera tubes is the difficulty of erasing completely a stored pattern of charge due to a 'hysteresis' effect in the photo-target. Unless special precautions are taken there is a considerable degree of carry-over between consecutive exposures with consequent loss of quality. This disadvantage of the vidicon tube is much reduced in comparable tubes using different target materials currently under development. (Plumbicon, Selenicon.) It is not practical to use vidicon, or any tube with storage, in irregular format searching scans since the stored charge is read out in an irregular manner and selective (partial) erasure of the stored pattern results.

Another method of use is to scan with a single vertical line and use the document movement for the horizontal scan component. In this way it is not necessary to 'stop' the document either physically or by using a short duration flash.

Scanning speeds with camera tubes are limited by the attainable deflection waveforms and/or usable signal to noise ratio obtainable—the latter is a fundamental and the former a practical limitation. Speed comparable with the flying spot scan might be achievable. The upper limit of resolution obtainable with available vidicon tubes is rather less than can be obtained with cathode ray tubes specially developed for use in flying spot scanners.

Photo-cell matrix (retina) Fig. 29

In contrast to the various scanning actions described above some use has been made of a fixed two dimensional array or matrix of photo-cells thus eliminating the need for any motion either of the document or within the scanning device. Since the reflectance at all points in the input character is measured in parallel (simultaneously) rather than sequentially the time to 'scan' a character is limited only by the response time of the photo-cells which can be reduced to fractions of a micro-second if need be.

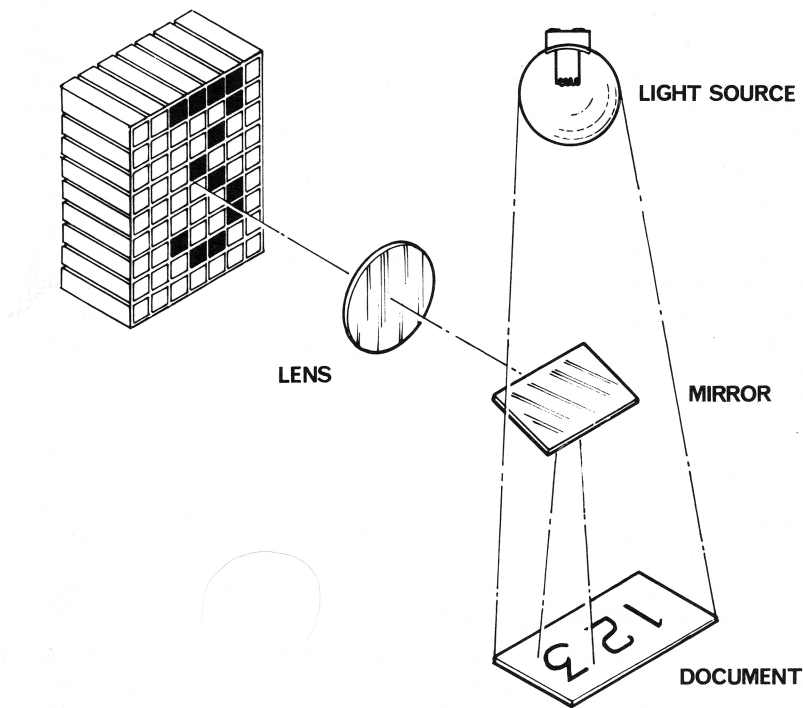


Fig. 29 A photo-cell matrix or retina.

The subsequent processing units are usually connected directly to the individual photo-cell outputs with the result that it is not economically possible to compensate for misregistration of characters falling upon the matrix. Mechanisms therefore have to be employed which sense the position of individual characters before projection onto the matrix. The necessary pre-scan can be performed by a second matrix of photo-cells suitably interconnected to detect the position of incoming characters the output of which is used to control positioning devices before presen-

tation of the character to the main matrix. Correction of position may be accomplished by appropriate presentation of the document or adjustment of the attitude of mirrors used in the optical system projecting characters onto the matrix.

Mechanical movement of the document and correct positioning of the image on the main matrix defines the system speeds obtainable with a matrix scanner.

Comment on scanning methods

Limitations of the single dimensional slit and divided slit scan systems are fairly obvious and were described under those headings. It is useful, however, to compare the two dimensional scanning methods. Of these the simplest is the mechanical scan using a rotating disc. This has the advantage of robustness and permits the use of abundant light upon the document to be read. The effects of ambient illumination, of moderate level, upon the scanner performance is then negligible and it is unnecessary for the reading station to be enclosed within a light tight box. It is also possible to use a coloured source of illumination with the consequent advantage of choosing a colour best suited to the document and ink colours to be encountered. The same comment applies to the slit scan methods in which the use of an infra-red sensitive scanner overcomes some of the problems of over writing with certain ink pigments if a true black printing ink is used for the characters to be recognised.

The flying spot scanner can achieve the highest resolution and hence largest field of view, without mechanical movement of optical components, of any type of scanner. It is capable of scanning at rates of several thousand characters a second but the choice of the colour of the light which can be used is limited. At highest scan rates the light obtained from a cathode ray tube is towards the ultra violet end of the visible spectrum. The presence of near u.v. will exclude the use of certain coated papers and ink pigments which fluoresce when irradiated at these wavelengths. It is also worth noting that the direction of illumination of characters with a flying spot scanner is highly constrained, being essentially at right angles to the paper surface. To avoid troubles due to texture shadows at the paper surface under these conditions the reflected light collection system containing the photo-cell must be non-directional in order to capture light reflected in any direction—this is conveniently achieved using an integrating light collection system, e.g. a photometric sphere, light guides, etc.

Vidicon t.v. camera tubes are small in themselves and use small optical systems resulting in a compact scanner. Using the single scan technique it is possible to read different size fonts merely by altering the length of line scan. Alternatively a limited amount of searching may be carried out by shifting the beam up or down. With an independent light source the colour of the illuminating source can be chosen at will subject to limitation imposed by the photo-conductive target. With flash tube working, exposure of the photo-cathode must be very brief, of the order of 100 micro-seconds in order to avoid the image stored by the tube becoming smeared due to significant document motion during exposure. The photo-cell matrix has no inherent storage and can use continuous illumination of the documents of any desired colour. Though rather bulky the matrix responds to the complete incident character

sensibly instantaneously with the possible advantage that many attempts to recognise individual characters can be made as optical components in the projection system adjust the position of the image on the matrix without unduly slowing down the overall recognition rate which will probably be limited only by the speed of paper motion. In addition it is not necessary to digitise the output of the matrix so that the grey scale information of the characters can be used.

This concludes the review of scanning methods used. The most common methods in practice are the mechanical scan (by far) and the C.R.T. flying spot scanner. Possible developments of scanning techniques will be described briefly in a later section.

Normalisation

The object of processing a character between scanning and attempting to analyse it and measure its properties is to convert the incoming raw signal into a (hopefully) standard form. Ideally characters would be presented to the property detection unit as correctly positioned black/white (binary) signals of fixed line thickness with no interfering noise or mutilations. Needless to say the ideal is rarely if ever achieved. In general the process of normalising requires some prior information about the incoming character, e.g. its range of ink density, position, etc.—since no such information is available until the character has been at least partially scanned a 'preview' of the character signal is essential.

Three methods of 'preview' have been employed—

- (a) a complete prescan of the character, in which the whole character is scanned in order to determine the range of ink densities and/or its exact position,
- (b) a partial scan—often a fraction of a line scan, from which the range of ink densities can be estimated and/or the 'leading edge' of a character detected and
- (c) to use information gained during the scan of the previous character which may be adequate for the succeeding one which, if a reject is detected and a rescan of the rejected character is initiated, the previous scan is in fact a prescan of the doubtful character itself.

Previewing is not usually employed in conjunction with slit scans in MICR or OCR systems and explicit normalising is not then possible in these systems but is implicit either in the maintenance of high print quality standards or the tolerance of subsequent processes to variations in the character.

It is common with systems using complete two dimensional scans to convert the one dimensional signal from the scanner back into two dimensional form for subsequent property detection by storing it as a pattern of 1's and 0's in, say, a two dimensional shift register (matrix of memory cells). In order to convert the continuous analogue output from the scanner to a binary form a threshold must be set at or above which 1 is recorded indicating 'black' and below which 0 is recorded indicating 'white'. This threshold should clearly relate to the actual (absolute) 'black' and 'white' levels existing in the character scanned. This information is obtained during the 'preview' of the character, the

best threshold setting determined and is then applied to the scanner output signal during the main part of the scan. The digitised signal is then fed into the shift register. Either by examination of the pattern of 1's and 0's in the register or during a second scan the position of the character in the scan is determined. It is then possible to shift the stored character into a preferred position in the shift register, or, if its position is determined during a pre-scan, it can be stored directly in the preferred position. The preferred position may be either with two extreme edges of the character abutting two edges of the shift register matrix or with the character centred on the matrix. It will be appreciated that the estimates of the character 'black' and 'white' levels and the estimation of the position of the character are critically dependent upon the print quality; distortion of character boundaries and/or excessive variation in character limb and paper density resulting in gross errors.

A normalising technique aimed at improving the quality of printed characters by removing in-filling within character loops, etc. has been employed in some readers. This improvement is achieved by effectively scanning the character with two coincident spots of different diameter, a small spot which scans the character at maximum resolution and a larger spot whose diameter may be an appreciable fraction of the character dimension, say several line thicknesses, Fig. 30. The small

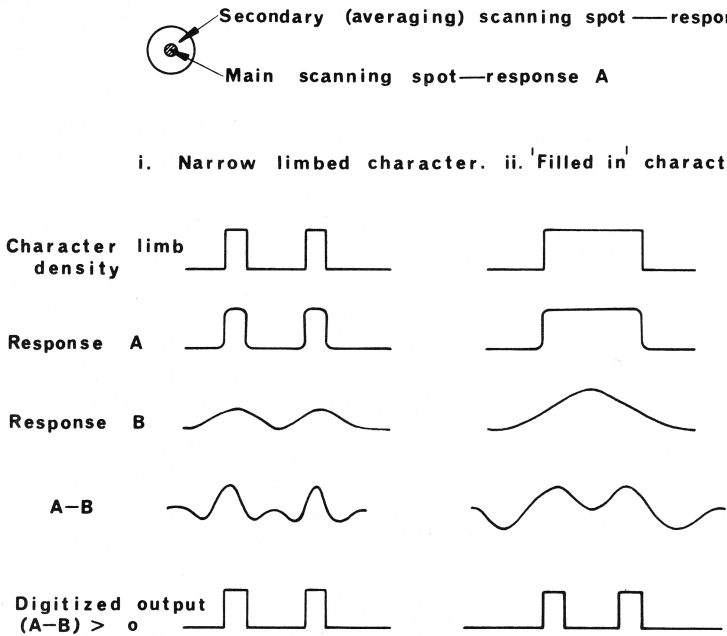


Fig. 30 Elimination of 'infall' by double spot processing

scan spot generates a high definition output waveform representing the character in the normal way, the secondary, large, spot generates an output proportional to the average reflectance of the character within the spot. Taking the difference (or ratio) of these two waveforms has the effect of suppressing signals from areas of constant density com-

parable in size with the large spot but leaves small details of the character resolved only by the small spot, unaltered. In this way severe infilling of character loops is practically eliminated. Additionally the binarising threshold required in such a scanning arrangement can be set permanently just above the zero level of the difference signal.

The double-sized spot scan has been achieved in practice by modulating the spot size of a flying spot scanner by alternately focussing and grossly defocussing; the time difference between the two spots being made by delaying one an appropriate amount relative to the other.

It is worth noticing that there are two distinct types of pre-normalising processes described above. The earlier processes for binarising and positioning were concerned with processing the character in such a way that the subsequent recognition equipment could operate at all effectively and inevitably result in further degradation of the character due to their ad hoc nature. The last process described was aimed explicitly at improving the character quality. It is probable that more processes of this nature will be introduced in future devices.

Character property detection

The purpose of the property detection unit is to extract from the signal representing input characters a set of characteristic measures or statements upon which a decision upon their most probable identity can be taken. The type and sophistication of properties detected depends upon the size of the set of characters to be recognised and the quality of the signal to be processed—which is itself dependent upon the effectiveness of the normalising operation and the print quality.

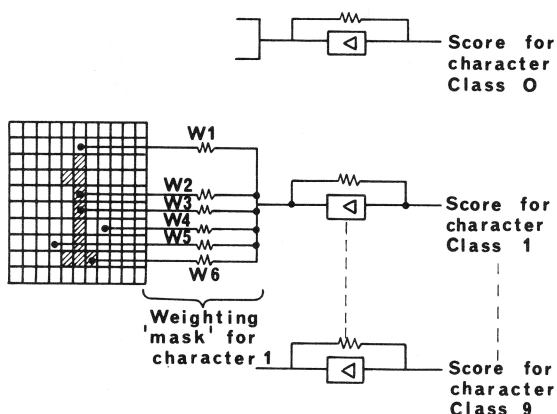
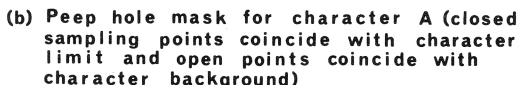
Mask matching

This is the most obvious property measure—the correlation or match between a specimen character and each of a set of masks representing the set of characters to be recognised. There are several forms of mask matching procedures.

Holistic masks—in which the complete character is matched against replicas of the theoretical character shapes, i.e. a perfect specimen. This is straight forward cross-correlation without any attempt to differentially 'weight up' those parts of a character which materially assist discrimination between character classes, (Fig. 31a).

Peep-hole masks—in which only selected areas in the character space are used as a method of comparing the incoming character with a set of standard characters, (Fig. 31b). The areas selected for use in this way are chosen on the basis of their uniqueness to particular character classes and are in effect highly weighted for this property, all other possible points of comparison being discarded.

Weighted area mask—in which all points in the character space are used in the correlation between unknown specimen character and the set of standards but are weighted on a continuous probability basis (rather than the discrete weights in peep-hole mask) according to their effectiveness in the discrimination between character classes. (Fig. 31c.)



(c) **Weighted area mask.**

Fig.31 Mask matching techniques

Comment on mask matching systems

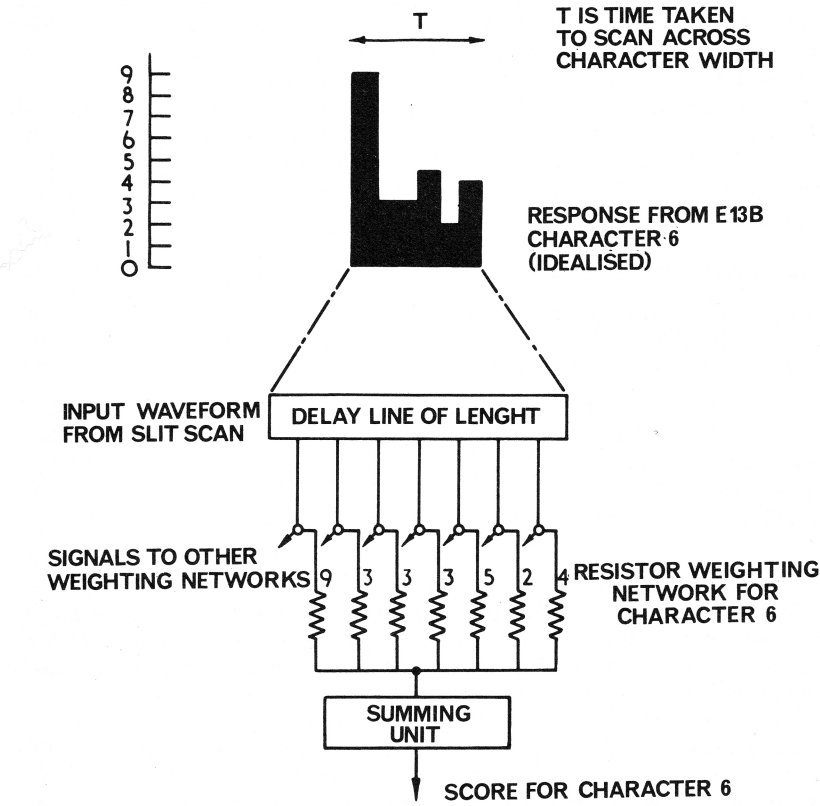
All direct matching systems of this general sort are critically dependent upon the position of the character as presented to the masks and the quality of printing. Holistic masks fail to make use of the obvious fact that some areas of any character are common to many character classes and conversely that some other areas are nearly unique to specific characters. Clearly best use of the available information is not made—this particular type of system although being developed comparatively recently has not appeared in a production machine.

Peep-hole mask techniques which have formed the basis of several commercial equipments acknowledge the varying significance of different parts of the characters in distinguishing between them, but on an all or nothing basis (binary weighting) using only a small number of the available comparison points. As a consequence they are rather sensitive to granular noise in the character i.e. small spots and voids. Weighted area masks make the best possible use of the available information and have formed the basis of several available machines and are being developed for some machines currently in pre-production state.

The decision mechanism for use with mask matching techniques has to do little more than indicate the best fit occurring between the unknown specimen and the set of standard masks. Some systems 'joggle' the representation of the unknown character into a number of small displacements from its standardised position in order to obtain the best possible fit and thus eliminate small errors in the character positioning operation. The decision mechanism also has to detect an ambiguous match in which two or more masks fit more or less equally well, in this case the character would be rejected. It is also possible to reject a character on the basis of a poor and hence unreliable match with any of the masks. The degree of match which should, theoretically, be obtained can sometimes be predicted from the input character itself. Under such circumstances some fraction of this theoretical maximum correlation value may be used as a confidence threshold. Any character failing to give a response greater than this threshold level is then rejected.

Matched filter Fig.32

The matched filter system is in effect a variation on the direct matching schemes in which the characters do not have to be deliberately



MATCHED FILTER-HIGHLY SCHEMATIC

Fig.32

registered upon the reference masks but are moved through all possible locations with respect to the mask and must clearly pass through the position of best match.

A practical realisation of this system is shown in Fig. 5 in which the output waveform of a scanner is continuously compared with a set of representations of standard waveforms. This comparison has to be undertaken over a length of signal sufficient to include the signal from a complete character. In order that the whole length of signal is available simultaneously to the comparison networks the scanner output is fed into a delay line of the appropriate length which is tapped at a large number of points along its length. The output at these points represents the complete signal obtained while scanning a space large enough to contain a complete character. This waveform is then matched against standard waveforms represented by resistor networks, one for each character class, attached to the delay line tapping points. The maximum response from a resistor network indicates both the most probable identity of the character scanned and that the character is correctly registered. It is assumed that the response is only maximal when exact registration occurs, but in practice the output from the networks is irregular and will probably have more than one maximum due to partial matches between a grossly mis-registered character, or perhaps the left and right halves of adjacent characters, and one or other of the resistor networks. It is therefore essential to estimate the approximate instant at which the maximum match is likely to occur. This is conventionally achieved either by detecting the blank space between adjacent characters or by detecting the total amount of signal in the delay line which will be a maximum only when it contains a complete character.

Comment on matched filter

This system is clearly insensitive to the exact location of characters due to the dynamic nature of the decision and it can be used to read characters of variable and irregular pitch.

Matched filters in conjunction with a slit scan have been used extensively in MICR systems using E13B character style. Extended forms of matched filter in which the delay line is replaced by a shift register are also being used in systems for reading ISO B character styles in conjunction with flying spot and vidicon scanners.

Stroke analysis Fig. 33

This general technique is distinguished from earlier masking operations in that recognition of specimen characters is based not on the correlation of the whole character with some standard, possibly deliberately weighted, reference masks but upon the assessment of a group of characteristic properties or features which describe each possible character class. In this way the recognition device is presented with a 'coded' representation of the original character. It follows that some features or code elements may be common to two or more character classes.

Ideally each code element or feature belongs to half the number of class and each feature to a different half. Characters have in fact been designed for use with the technique which differ in at least two code

elements. The character style ISO A (OCRA) recently adopted as an international standard type face for machine reading is intended to be recognisable simply from a statement of the relative positions and lengths of vertical and horizontal straight line features which it contains—other details being introduced largely to improve the legibility of the characters for human readers. Other styles depending upon the same general principles (dual coding) are N.C.R. Bicode, S.E.L. CZ13 and Farrington Selfchek.

The features to be detected in these coded characters are simple and may, if the characters are well printed, be detectable in the output of a

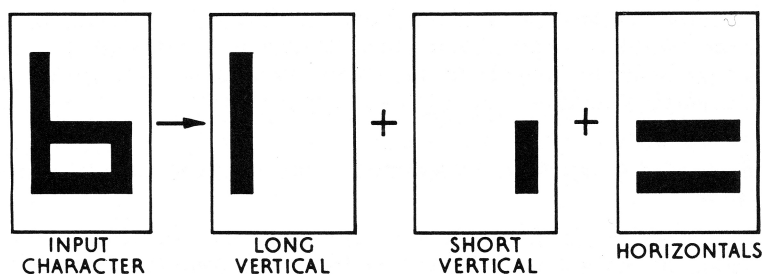


Fig. 33 Stroke analysis

vertical divided strip scan of low resolution. A vertical feature can be characterised by simultaneous activity in several, say three, adjacent cells in the scan. Horizontal features are detectable by continuous activity in isolated single or adjacent cells of the divided strip.

The sequence and relative positions of detected features is the used to identify the characters.

Comment on stroke analysis

Stroke analysis is limited in application to character styles specifically designed for this type of analysis e.g. ISO Style A numerals. The range of clearly discriminable symbols which can be generated for this method of analysis is limited—probably to the numerals plus a few control symbols. However if the range of strokes to be detected can be extended to include diagonals of varying length and in both principal directions the repertoire can be extended to include alphabets if well printed, but with considerable increase in the complexity of the reader.

Decision mechanism

The decision mechanism has to indicate the most probable identity of the unknown character upon assessment of the information gained in analysis. In equipment currently available employing mask matching or match filter techniques the decision is based upon a comparatively simple test to determine which one of a set of properties is the largest. In the stroke analysis system the presence or absence of particular

strokes is indicated by a binary (two state) variable and the decision is based upon the comparison of the detected states of the possible stroke positions and each one of a set of standard statements stored in the machine.

The method of stroke analysis and subsequent recognition of the unknown character by the detection of a combination of properties (strokes) is a simple example of more elaborate ~~two step~~ processes currently under development for more elaborate CR situations when control of type style and/or printing quality is not possible to any great extent. These systems will be described briefly in a later section of this chapter.

In addition to assessing the most probable identity of an unknown character the decision mechanism must also assess the reliability of its decision. Should the reliability fall below some prescribed limit then the character should be classified as unrecognisable by the system i.e. be rejected.

In the event of a reject being detected several courses of action are available

- (a) to simply reject the whole document possibly indicating by an index mark the approximate location of the doubtful character whose identity can then be indicated and inserted by the system using a correction list or by complete manual transcription of the rejected document,
- (b) to reread the line containing the reject character in some cases under changed reading condition,
- (c) halt the reading process and document motion and indicate, with a visual display, the reject character for operator action—key in identity,
- (d) display of the reject character on a storage display device for action by the operator. Correct identity of the character is then automatically inserted at the correct place by the system. The system continues to read having displayed the reject—until a second reject occurs or until all the available storage displays available are committed when the system must stop.

Output

Recording the output from CR equipment poses no special problems beyond providing a recording means compatible in speed with the reader—this often excluded punched paper tape and probably punched cards which are limited to speeds of the order of 100-300 characters per second. Magnetic tape can clearly be used as an output medium or direct on-line working into a computer is possible. The latter possibility permits reading-format control to be exercised by the computer which could effectively speed up the effective reading rate by, say, skipping over irrelevant areas of a document.

Future developments and prospects

New ideas are continuously emerging in the field of OCR and it is not possible to predict with any certainty what the next stages in development of techniques will be or when new systems will appear commer-

cially. However it does not require much imagination to see in what respects existing techniques should be improved and it may be useful to explore some of the ways of obtaining such improvement currently being developed.

The basic criticisms of available machines centre around the cost of installing such systems (including the cost of modifying or replacing existing printing devices if necessary), the error and reject rates obtained in practice and the requirement to print to a high standard in a single type style. This latter objection may be alleviated to some extent as the proposed international type faces are in fact adopted. However the adoption of international type font standards does not guarantee their immediate widespread application, and the anticipated desire to use CR systems for reading documents from a variety of uncontrolled sources will emphasise the need for multifont readers capable of working with a realistic, i.e. conventional, range of print quality.

Devices and techniques described in this section are not yet available in production equipment, indeed many are still speculative and unproven research and development ideas and may well be discarded eventually.

Scanning techniques

Scanning techniques are unlikely to change dramatically in the foreseeable future. Assuming that the demand for high throughput systems continues, mechanical scanners will inevitably give way to flyingspot or t.v. camera types. T.v. scanners have the advantage of using a preferred source of illumination colour and independence of ambient light conditions, but the inability to 'format search' with storage tubes as currently used will limit their use. A non-storage type of camera tube which was developed in the early days of t.v., the image dissector, might find a new lease of life in this application. Development of the image dissector could lead to the flexibility of scanning obtained with a flyingspot scanner combined with the facility for using optimum document illumination conditions of the t.v. camera, and comparable resolution and scanning rates should be achievable. A new non-storage t.v. camera device, recently demonstrated, is the 'scanistor'. This is a wholly solid state image dissecting device, which in the current state of development performs a one dimensional scan only. It is effectively a single column of photocells fabricated in such a way that each element in the column can be interrogated in turn by the use of a single analogue 'scanning' signal to produce a time varying analogue output signal from the device. A two dimensional solid state scanner working on the same principle may be developed. It is too early yet to say what resolution or operating speeds will be achieved with the 'scanistor'.

Normalising

With the development of new property detection techniques (q.v.) which are less critical on the location of specimen characters the demands on position normalising systems will decrease with a consequent reduction in the cost of equipment for this purpose.

A normalising technique for improving the quality of a printed character has already been described. Other techniques for performing similar and allied processes are known to be under development. In particular the problem of detecting the middle of character limbs so

that presented characters can be reduced to standard width forms are being studied.

Property detection

The idea, introduced in describing the stroke analysis method of property detection of detecting a number of properties e.g. the existence of strokes in certain positions in the character, which in combination serve to identify the unknown character is a recurrent theme in several CR schemes in various stages of development. The general thesis is that a number of limited properties of a character, e.g. intersections, corners, curves, straight lines etc. are individually more reliable than the whole character and lead, in combination, to a more reliable decision on the most likely identity of a character than the global matching techniques in which these various features are implicit but cannot be separately 'addressed'. In a practical system the property detectors determine the existence of some characteristic shapes and their relative positions. The properties may be defined by operators containing n rigidly connected sampling points, n -tuples, which are exhaustively scanned over the character by some means which will not be discussed here. During this scan each n -tuple will 'fit' the character several times if it contains the corresponding property. The n sampling points of each n -tuple are associated by some logical or arithmetic function, the sense of each point determines whether it responds to the character limb or the character background. Both response senses are necessary in order to discriminate between say, the open loops of a 3 and the closed loops of an 8. The property measures passed on to the decision mechanism may be either the (binary) statement that the n -tuple has or has not fitted the character at least once or a parameter indicating the aggregate fit between the n -tuple and the character. The latter would appear to give more information than the former.

This method of character analysis is fundamentally invariant to the exact position of the character within the field scanned since the operators are evaluated at all possible positions within that field. The position normalising operations can be considerably simplified as exact location upon a fixed mask is no longer necessary. Systems in which the n sampling points are associated by an arithmetic operation (multiplication) use the analogue video signal representing the character rather than the discrete two state derived waveform. Hence the equipment to transform the continuous to a two state signal is not required. As observed earlier this transformation is somewhat arbitrary and tends to further degrade character quality, it is possible therefore that the use of the continuous signal apart from simplifying the normalisation process might also result in an increased print quality tolerance.

Systems using the 'scanning n -tuple' principles are under development in U.S. and Britain for use in multi-font and low print quality situations. An approach to the problem of multi-font recognition which has appeared in U.S.A. is to identify the type style before attempting to recognise individual characters. The only constraint upon the production of documents for this reader, apart from the maintenance of a good quality of printing, is that the first character following a change in style must be upper case.

Decision mechanisms

To match the increased sophistication of property detection methods decision mechanisms are also being improved. As was noted in discussing stroke analysis techniques decision may be based upon the existence in a character of a combination of properties. In the earlier instance the combination tested is the logical AND of a strictly limited number of properties (character strokes); but this arrangement does not allow any latitude in the decision—it is either exactly correct or, if any single stroke is missing, wrong. If the number of properties detected is increased above the minimal number of information bearing elements used in stroke analysis then the characters are over specified i.e. the property measures form a redundant description of the character. In this situation exact match of a combination of properties with a standard set is not essential and some distortion of the property measures is permitted without risk of mis-recognising the source character. It is not possible to describe proposed decision mechanisms here in any detail, suffice to say that many of the more practical systems employ linear decision functions which in essence form linear weighted sums of the property measures. These sums or 'scores', one or more for each possible character class, are either compared with predetermined thresholds or against one another to detect the maximum 'score' which indicates the most probable identity of the input character. The decision mechanism is the only section of the general character recognition system which has so far proved amenable to mathematical abstraction—processes of normalising and property detection being largely a product of the intuition and insight of individual workers, no general theory has yet emerged for either process. Readers who wish to obtain more than the very cursory discussion given above are therefore referred to the literature (cf Sebestyen 1962).

Further departures

Beyond the natural progression towards the development of readers for the recognition of multi-font, low print-quality material and cheaper reading devices there is a considerable amount of research under way into the problems of recognising hand (block) printed numerals both in U.S. and Europe.

Further on still is the possibility of recognising cursive script—this possibility is remote at present since context plays a large if not decisive part in reading script. Computers are not yet available which can be used economically or at useful speeds for performing the elaborate dictionary searches and syntactic analyses necessary when making use of context. There is serious doubt whether the reliable recognition of script will ever be a commercial proposition.

References

- Stevens, M. E. 'Automatic Character Recognition—a state of the art report'. M National Bureau of Standards. Washington. May 1961.
- Bijleveld, W. J. 'Automatic Reading of Digits'. The Netherland Automatic Information Processing Centre. Amsterdam. March 1963.
- Sebestyen, G. 'Decision Making Processes in Pattern Recognition'. Macmillan Company. 1962.

SECTION 6

Proprietary Equipment and Services

Opinions expressed in this chapter are those of equipment manufacturers and suppliers and are not necessarily endorsed by The British Computer Society.

The ODP system

The design and development of character recognition equipment has resulted in machines of considerable technical sophistication. But genuine admiration of the latter should not cause us to lose sight of the prime purpose of Character Recognition: to provide automatic input to a computer from characters which are also visually legible to laymen. There is, of course, implicit in this the complementary requirement that the initial creation of the characters shall be simple and economic and to the required level of accuracy. Since it satisfies both parts of the requirement, Cummins ODP (Original Document Processing) system constitutes another form of Character Recognition which already has eleven users in U.K. covering finance houses, utilities, multiple re-tailing and local government.

Unlike MICR and OCR, in which the characters are imprinted on the paper, ODP digits are perforated through the paper, Fig. 34. In many aspects of data processing all or part of the data for an updating entry

BRANCH	DUE DATE	AGREEMENT NUMBER	INSTALMENT			LLOYDS BANK LIMITED CREDIT TRANSFER																																					
421204567			£	s.	d.																																						
30-00-00			Lloyds Bank Limited HEAD OFFICE COLLECTION A/c No. 25-01 ANGLO AUTO FINANCE CO. LTD. 1 A/c			DATE 18.7.66																																					
Cashier's stamp and Initials						<table border="1"> <tr> <td>£5 and £10 NOTES</td> <td>30</td> <td>0</td> <td>0</td> </tr> <tr> <td>£1 NOTES</td> <td>5</td> <td>0</td> <td>0</td> </tr> <tr> <td>10/- NOTES</td> <td></td> <td></td> <td></td> </tr> <tr> <td>SILVER</td> <td></td> <td>18</td> <td>6</td> </tr> <tr> <td>BRONZE</td> <td></td> <td></td> <td></td> </tr> <tr> <td>COPPER</td> <td></td> <td></td> <td>3</td> </tr> <tr> <td>TOTAL OF CASH</td> <td>35</td> <td>18</td> <td>9</td> </tr> <tr> <td>CHEQUES, P.O.'s, etc.</td> <td></td> <td></td> <td></td> </tr> <tr> <td colspan="4">£ 35 18 9</td> </tr> </table>		£5 and £10 NOTES	30	0	0	£1 NOTES	5	0	0	10/- NOTES				SILVER		18	6	BRONZE				COPPER			3	TOTAL OF CASH	35	18	9	CHEQUES, P.O.'s, etc.				£ 35 18 9			
£5 and £10 NOTES	30	0	0																																								
£1 NOTES	5	0	0																																								
10/- NOTES																																											
SILVER		18	6																																								
BRONZE																																											
COPPER			3																																								
TOTAL OF CASH	35	18	9																																								
CHEQUES, P.O.'s, etc.																																											
£ 35 18 9																																											
PAID IN BY <i>R. V. Pinner</i>																																											
⑈ 30 0000 ⑈ 0250 104 ⑈																																											

Fig. 34

is known in advance; what is significant is the time and, sometimes, the place at which that data re-enters the system. Instalment payments are an obvious instance of this but there are many others e.g. in applications covering house rents, rates and works orders. The particular merit of perforation, as a character recognition medium, lies in its inherent 'ganging' facility since identical data is entered into as many as fifteen sheets of paper in a single operation; it is standard practice for ODP users to produce all the updating data for the duration of a hire purchase contract in little over half a minute.

This method of producing instalment posting slips was in use long before the advent of EDP; the perforations, forming the outline of numerals, had been found to be wholly acceptable and intelligible to laymen and accounting machine operators alike. So, once computers came on

the scene, it was logical to develop the means of reading the perforations by machine to produce automatic input.

Like CMC 7 the ODP digits are coded; only a limited part of the digit is sensed both by the reading mechanism and the human eye; the rest of the digit outline is ignored by the machine and is there solely for human reading. Four binary 'bits' are sensed from the digit outline while a fifth, parity bit is sensed beneath the digit. As a result, a very high reading accuracy is obtained economically since the obscuring of a hole or the presence of a large pin or staple hole (in the limited sensing area) causes the reading circuits to reject the digit. In addition to perforating the digit outline and the parity hole, when appropriate, ODP perforators also produce a larger, sprocket hole below each digit.

The paper transport of ODP readers and sorters (Figs. 35 and 36) feeds each document, in turn, from the feed hopper to a revolving drum. The latter has sprockets on its circumference on which the document's sprocket holes are located. The document passes between a light source and a reading head containing a number of photo-electric cells. As perforated holes pass in front of the photo-cells, the latter generate pulses which are routed to a series of gates; these are opened, once for each digit, in synchrony with the sprockets, allowing the code pulses to reach the matrix decoding stage. Thus the additional holes, needed by the human eye, are diverted out of the machine logic. In their basic reading method there is a clear similarity between ODP and photo-electric paper tape readers; as a result, they are both typified by high accuracy combined with relatively low cost.

In common with all document readers, as opposed to page readers, much of the complexity derives from the handling of discrete pieces of paper. Unlike punched cards, the forms are relatively insubstantial and require handling with mechanical care. Although the sprocket holes have to be accurately aligned to the drum sprockets, the alignment mechanism of Cummins transports is such that documents do not have to be carefully jogged before they are put into the machine. The faster range of machines can even accept mixed sizes of documents in a batch. It is, perhaps, relevant to mention here that the first production OCR 'A' sorter-reader, with optical logic made by Standard Elektrik Lorenz, incorporates Cummins ODP paper transport.

Having decoded a digit, the ODP reader, acting through appropriate shift registers, operates an associated recording device i.e. a card or paper tape punch or magnetic tape unit. With the exception of readers linked to magnetic tape, it is the speed of the associated punch which determines the operating speed of the ODP reader; with a Teletype 110 chrs/sec paper tape punch, the Model 227 ODP reader reads 150 to 225 documents per minute; the throughput is, as usual, inversely proportional to document length.

It was shown earlier that ODP reading is parity-based so that a damaged digit is rejected. The reading system is programmed to accept only a specified number of digits per document in any one batch. The correctly-read digits from the document are counted electronically. Provided the number of digits is correct at the end of reading, the document is stacked in the 'read' bin and a signal is given to the recording device e.g. 'end of word' code in the case of paper tape. If the digit-count is wrong, the document goes to the 'non-read' bin and an 'erase'



Fig. 35

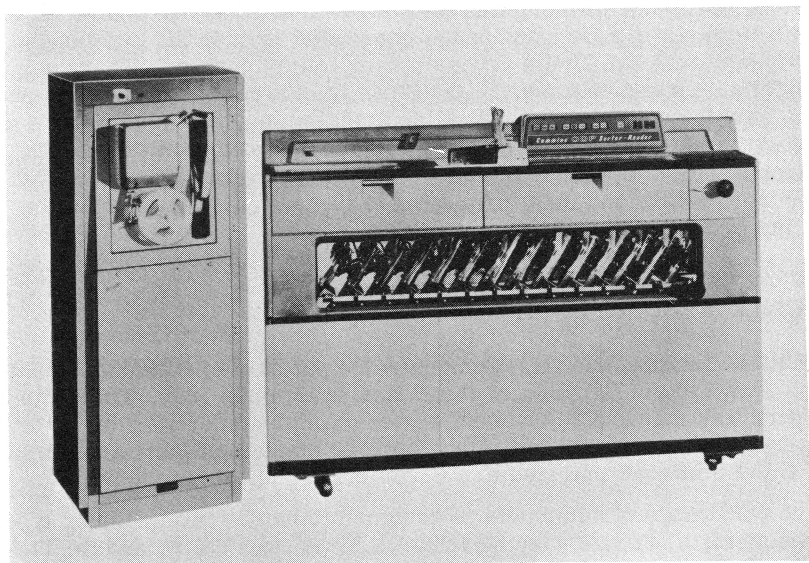


Fig. 36

signal is emitted to the recording device. The system is on a 'fail safe' basis so that a machine fault also results in rejection. The different counting programmes offered are called by operating a switch at the start of a run. The programming varies with the individual application. In one installation for example, only a part of the document is read on the first run, so as to prove the amount totals to the bank batch credit; the documents are then sorted off-line prior to a second run in which all the data is read for computer input. In addition to the data checking provisions, ODP readers and sorters have safety devices to protect the documents themselves. The thickness of the document emerging from the feed hopper is sensed; should two documents pass the mechanical separator, due to maladjustment, the double thickness causes the transport to stop and an explanatory signal to be lit. Similarly, a document jam, which occurs with ODP as with other machines, results in an immediate stopping of the mechanism; as a result the documents are readily removed, unfolded and returned to the feed hopper. Users' experience is that, provided the correct quality of paper is used, there is a negligible incidence of unprocessable documents; some documents being read to-day were supplied to H.P. hirers four years ago.

Character Recognition frequently calls for two types of encoding; pre-coding e.g. serial number, bank and branch code and customer account number on a cheque, and post-encoding e.g. transaction code and amount; the latter, of course, occurs after the document has re-entered the system and before it is read. With ODP, post-encoding is carried out either on-line, in low volume cases, or off-line with an amount encoder. In the on-line case the reading of a document is interrupted so that additional or substitute data can be visually read from the document and keyed directly into the associated recording device. In the off-line case, the encoder is used to perforate the additional data into the document itself; consequently, reading speed is not lowered. The encoder comprises an add-lister controlling a perforator; the latter aligns the document by its sprocket holes so that the additional digits are perforated in prolongation of the original code line and in register with it; this is the purpose of the unencoded area to the right of the code line on most ODP documents seen in this country. Where, e.g. in instalment work, an amount, other than the correct instalment, is paid the encoder not only perforates the actual amount but also introduces a programming instruction into the document; as a result the reader automatically suppresses the reading of the original amount so that the word length and data format fed to the computer is constant for both regular and irregular documents.

OCR 'A' READING EQUIPMENT

OCR 'A' reading was first demonstrated publicly by Automatic Input Systems Ltd., of Croydon, at the B.E.E. in October 1965. The equipment shown was the ODS 2 document sorter-reader comprising the optical system and electronic logic developed by Standard Elektrik Lorenz A.G., of Stuttgart, married to a Cummins document transport.

The ODS range of equipment is designed to handle two differing types of document; (a) Forms varying from $2\frac{1}{2}" \times 4"$ to $4" \times 9"$ and (b) Tickets from $0.8" \times 2"$ to $2\frac{1}{2}" \times 4"$. Each document can have a single or double code line located anywhere on the document apart from a $\frac{1}{4}"$ margin at

the edge. Any character density up to 10 to the inch is acceptable so that ODS is compatible with standard computer print. A mark-reading option is an available alternative to the second code-line.

SEL's approach to machine reading renders its reading equipment remarkably impervious to dirt, rubber stamping and other extraneous marks. This is because it has been designed to respond to infra-red reflections of the scanned image. As a result, marks, which, to the human eye, appear so dark as to impede the reading of the print beneath, are not 'seen' by the ODS reading head which, therefore, reads the data with accuracy. Many exhibition visitors, seeing the ODS 2 read, without error, digits so obscured by blue biro as to be visually undecipherable, felt the machine must be using magnetic sensing.

Complete reading logic is contained in each machine; they are, therefore, available for use either on-line or intercoupled to a magnetic tape unit. ODL3, the ticket machine, is a reader only. ODS2, the sorter-reader for cheque sized forms, is in production and ODL4, its three pocket version, will shortly become available.

The scope of the S.E.L. reading system has, for the present, been limited to the ten numerals and three operating symbols. There is, on the credit side of the stylisation of OCR 'A', a very wide separation of identity in format between each of the digits. Thus there is a difference of four shifts of the lower vertical between the identity of a figure two and a three; similarly, between a nought and an eight, there is a single shift of each of the upper verticals and the presence or absence of the middle horizontal. Prior to the introduction of OCR 'A', SEL did a lot of development with CZ13, its own OCR font; the latter, unlike O.C.R. 'A', did not utilise the horizontal strokes in machine reading. Yet an independent, governmental body, making a comparative test of character reading equipment, recorded CZ 13 as having an error rate of 0.25 per million characters read, the best result of all the equipments under review.

Like CZ 13, its forerunner, ODS/ODL employs a dynamic reading method. The documents move at constant velocity under a scanning slot. In the rare event that a document does not read satisfactorily it goes, on a 'fail safe' basis, to the 'non-read' bin. The maximum reading speed is 900 tickets per minute for ODL3 and 700 forms per minute for ODS2; these figures, as is conventional, relate to the minimum document length and are lower for longer documents.

For further information please contact:-

MR. A. V. J. BRENNAN
Grosvenor House
125 High St
Croydon Surrey

BUREAUPREP Ltd

Data Preparation & File Creation

by KEYPUNCHING and

OPTICAL CHARACTER RECOGNITION

30-31 THE BROADWAY · EALING · W5 · GEOgian 4343-4

2. CONTROL DATA LIMITED

Control Data 915 Page Reader

Introduction

Information in the form of numbers, letters, signs and symbols is the raw intelligence of data processing systems. Converting raw information into 'computer language' requires data conversion techniques which are expensive and time consuming.

The Control Data 915 Page Reader enables this expensive data conversion to be eliminated, by permitting direct translation of typewritten or printed information into computer language by optical electronic methods.

The 915 Page Reader

The 915 Page Reader (See Fig. 37) reads all letters of the alphabet, numerics 0-9, standard punctuation and special symbols. The Reader handles documents of all sizes from 4 to 12 inches in width, 2½ to 14 inches in length, and continuous stationery.



Fig. 37

The Page Reader operates in conjunction with a computer which directs its reading programs, and forwards the data to any computer-driven recording device, such as a magnetic tape unit, line printer or card punch.

The functional drawing at Fig. 38 shows the arrangements of the reading station and the document handling processes.

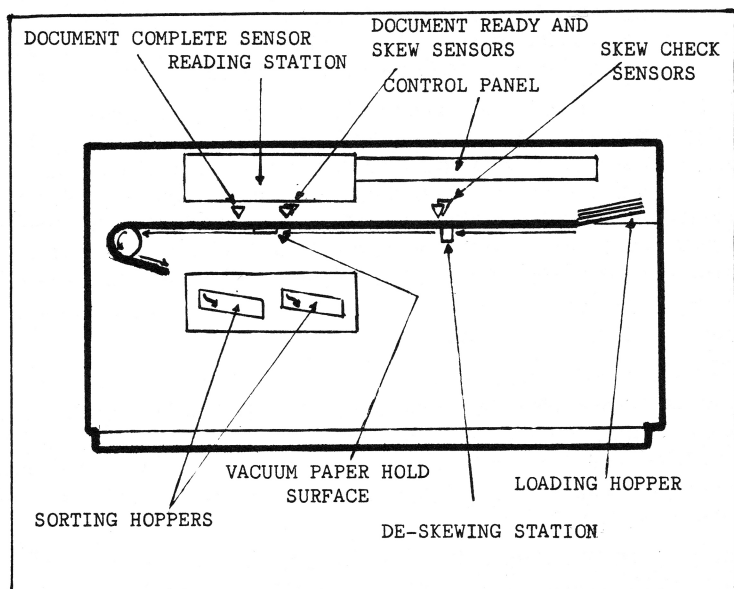


Fig. 38 Control Data 915 page reader;
Functional diagram

Document Handling

Documents are moved from the loading hopper, through the reading station, by vacuum belts which traverse the Reader's open table. The vacuum belts move in a stepping sequence of $\frac{1}{60}$ th or $\frac{1}{600}$ th of an inch per step, or a combination of both, as programmed. Movement from the reading station to the sort bins is by high speed friction belts, inter-clutched with the stepping belts to ensure a quick, continuous movement in a single cycle. Transfer from the reading station to the sort bins takes place in 250 milliseconds.

The Reading Cycle

When a document has arrived at the reading station, control is automatically transferred to the computer which then directs the reading of the document as well as the processing of the data to be read. Illuminated by powerful lights, character images are reflected by a mirror that is in a horizontal position above the reading station through a lens system to a set of light fibres, which, in turn, transmit the light or dark image elements to photo-multiplier tubes. Converted to electronic signals, the character elements are fed into quantizing circuits and loaded into shift registers, thence to analogue resistor matrices, comparator voltage selector circuits, and coding circuits, and are input to the computer in binary code through the Readers' controller.

Reading Speed

Because the Reader only reads that portion of a line which it is instructed to read, and does not waste time scanning blank lines and unwanted data, a very high speed of operation results. Specifically, direct optical to electronic conversion gives a line reading speed of 37 inches per second—or an instantaneous reading rate of 370 characters per second.

The Reader's rapid operation results in an outstanding reading capability. For example, a standard $8\frac{1}{2} \times 11$ inch document prepared in a six line per inch format with full line use and an overall margin of 1 inch can be read in fourteen seconds; a 3×5 inch document with a four line format four inches long can be read in $1\frac{1}{2}$ seconds. The reading time is reduced considerably when the text is 'unjustified'. The Reader recognizes odd line lengths and does not waste time scanning the unfilled remainder.

Checking

Two types of skewing problem are automatically handled by the Reader. The first is 'paper skew', which is checked at the skew stations. The second is 'line skew' in the type or print line of the document itself. This is compensated for by the Reader to up to one character height in six linear inches of line.

Jamming may result when a document is physically skewed out of tolerance. In this case the Reader stops operating, and an operator's indicator lights.

The 915 can read misaligned characters to a maximum of 1 character height in a 3 line per inch reading mode, $\frac{1}{4}$ character height in a 6 line per inch mode.

A marking pen at the document reading station may be operated under computer control to mark documents for reference.

Mark Sense

The Mark Sense feature allows the Reader to recognize filled or unfilled circles, and to take special discrete action as a result. For example, the mark sense mode may be used to count or to cause immediate sorting of the document being read.

Type Font

The type font to be read by the 915 Page Reader is the proposed Standard font of the American Standards Association for use with optical character readers. The font is an upper case, or capital letter size, and includes the 26 letters of the alphabet, 10 numerals, 18 standard symbols, 5 abstract special symbols and 4 unique information symbols.

Computer

The Reader operates in conjunction with a Control Data computer, which may be the small 8092 or 8090 computers, or more powerful computers in the Control Data 3000 series.

The computer governs, by program function codes, the document positioning, reading, and sorting operations of the Reader. Paper motion

and the optics system are directed by the program so that information is read on a line-for-line basis.

Through a Control Data Digital Communications Terminal the Reader can be linked with other computers and peripheral devices at remote locations.

Summary Specification

Document Handling Capabilities

Input Hopper	:	Adjustable width, 4 to 12 inches. Average load: 1 inch of documents.
Sort Bins	:	Maximum load: 2½ inches of documents. Adjustable width.
Document Size	:	Width: 4 to 12 inches Length: 2½ to 14 inches Continuous stationery.
Document Stock	:	Weight: maximum thickness: 007 inch minimum thickness: 0028 inch Rag content: zero content desirable
Document speed	:	Input hopper through reading station: stepping rate: 11 inches per sec. Reading station to sort bins: 250 msec.

Reading Capabilities

Instantaneous reading rate	:	370 characters/sec.
Forward reading speed	:	37 inches/sec (To allow for recovery time, optic mechanics, consider the print line to be 1 inch longer than actually on the document).
Reverse/retrace time	:	148 inches/sec (Make same calculation as above)
Maximum line sweep	:	11½ inches, programmed.
Maximum line density	:	6 lines per inch, programmed.
Compensation for vertically misaligned characters	:	Programmed: 1 char. height in 3 lines/inch mode ¼ char. height in 6 lines/inch mode Re-scan capability for misaligned, illegible or marginal characters.
Type font	:	A-Z, 0-9, punctuation and special symbols.

Operational Requirements

Weight	: 1500 lbs
Size	: Height: 57½ inches Length: 81½ inches Width: 29 inches.
Heat dissipation	: 8500 BTU/hr.
Environment	: Typical computer, Temperature 60-80°F. Relative Humidity 40 to 60%
Power	: 208 V. AC., 3-phase, with neutral 20 amp fuse/phase.
Computer	: Control Data 8092, 8090 or 3000 series.

For further information please contact:-

CONTROL DATA LTD
26-32 Caxton Street
London SW1

Telephone No. Sullivan 3701

3. CROSFIELD ELECTRONICS LIMITED

Introduction

The Data Processing Division of Crosfield Electronics Limited supplies a range of fast and inexpensive machines for reading or microfilming cheques, GIRO forms, lottery coupons, or other transaction documents. Reader operation can be 'on-line' to computers, if necessary via a data transmission link, or through some intermediate medium such as punch tape.

All of these machines, except the Ticket Reader, employ Crosfield vacuum-handling techniques, ensuring that documents can be handled thousands of times at high speed (e.g. 1, 200 per minute) without appreciable damage, and making possible the use of very cheap and thin paper, e.g. for lottery coupons. At no time do paper documents pass through narrow gaps: thus jams and wrecks are avoided. All machines are equipped to deal with faulty documents.

Also available are Bank Note Counters, and instruments for inspecting the quality of magnetic or optical characters.

Document Transports for Character Recognition

The series 3000 Document Transport can be fitted with one or more of the following facilities: -

- (a) magnetic ink character reading
- (b) document counting for document control
- (c) document cancellation, to prevent accidental re-processing of documents.

In some instances this Reader, together with a suitable interface, can be provided for 'on-line' computer operation at about £5, 500

Document Transports for Microfilming

Document Transports can be equipped with illumination, camera, and printing facilities so as to constitute a High Speed Microfilming System operating at 752 documents per minute. The documents are printed with a serial number immediately before the microfilming process: this printing appears on the film and constitutes a simple and reliable indexing system.

Reader-Sorters

A range of Readers with three or more output pockets is available fitted with various combinations of reading, counting, serial numbering and microfilming facilities. A typical machine is the 3-pocket Model 1200 illustrated (Fig. 39 and 40). This is equipped to feed 9" long documents, and to read CMC-7 characters. Sorting is normally under internal control in accordance with information read from the code line of the documents. Alternatively, sorting can be under external control



Fig. 39 Document reader and sorter Model 1200 three pocket

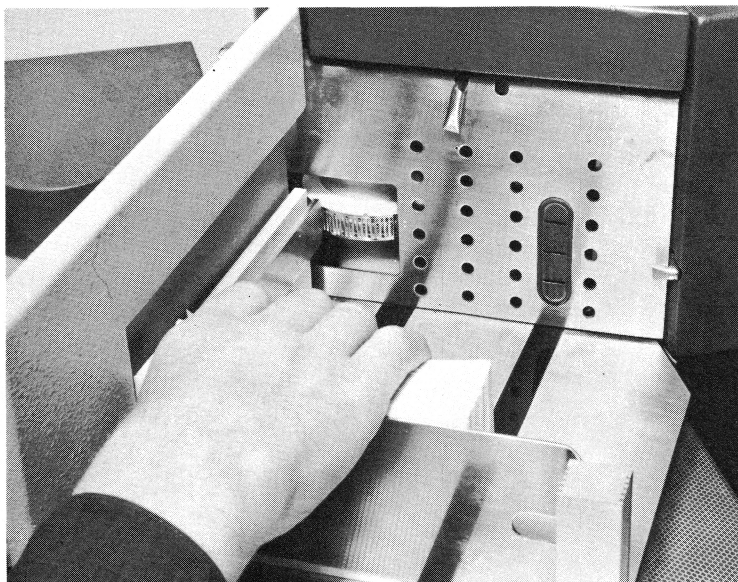


Fig. 40 Document reader and sorter Model 1200
View of input pocket showing peeler wheel and back post

from a computer. Some machines have been fitted with equipment for optical sensing of handwritten marks.

Ticket Readers

The Model 8000 Crosfield Ticket Reader was built for reading price tags attached to goods to be sold in retail premises. These tags carry price and product identification information in the form of CMC-7 characters. The tags are in multi-part form, and provide a convenient means of:-

- (a) recording details of goods transferred from warehouses
- (b) stocktaking
- (c) recording details of sales

Tickets are normally hand-fed. The output from the Reader can be taken to an accounting machine or via a data transmission system to a computer. Facilities are provided for dealing with damaged or defective tickets.

Bank Note Counter

The Crosfield Model 5200 Bank Note Counter is capable of handling bank notes or other documents, in both new and used condition, at the rate of twenty per second. Duplicated counters are provided, and a key-board type of batching control is available, for making up wage packets.

Bank Note Sorting

Bank Note Sorting machines are under development. It is envisaged that these will, in most instances, sort notes in accordance with coded magnetic marks, applied to the notes when they are first produced.

Micrex

The Crosfield Micrex testers are in use in most countries for checking the quality of magnetic characters, both E.13.B and CMC-7. The Table Model is particularly useful to printers, since it makes possible the examination of large sheets of uncut cheques. The Portable Micrex is

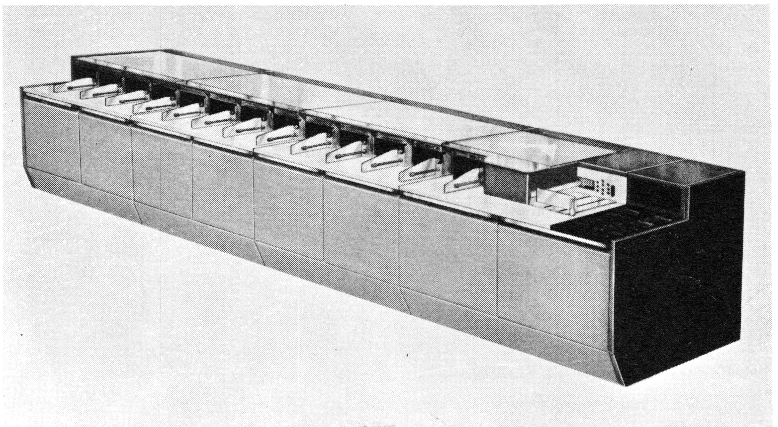


Fig. 41 Linear sorter. Thirteen pocket



Fig. 42 Document transport with CMC-7 read logic

a convenient instrument for use in data processing systems, handling magnetic coded documents. The Web Micrex is available for use on web fed printing presses.

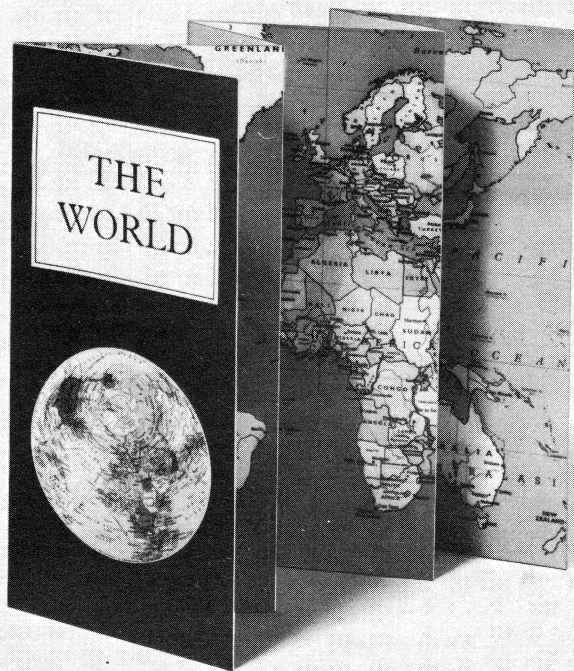
Optical Print Quality Monitor

This instrument is particularly suitable for use by printers in controlling the quality of printing destined to be subsequently read by an optical character reading machine.

For further information please contact: -

Mr C. H. Lees
CROSFIELD ELECTRONICS LTD
766 Holloway Road
London N19

**De La Rue Bull get all their computer
experience from just one special area . . .**



Here's a street guide

De La Rue Bull are part of the worldwide Bull General Electric group—whose data processing systems serve nearly 4,000 organisations in over 50 countries.

When you call De La Rue Bull, the group's resources are immediately yours. Factories and research centres in America, France, Italy and Holland can provide the answers to a very wide range of data processing problems.

When you need a computer, talk to De La Rue Bull *first*—or fill in the coupon.



DE LA RUE BULL

To: De La Rue Bull Machines Ltd.,
114-118 Southampton Row, London W.C.1.

I have some queries on data processing. Please contact me immediately.

NAME

POSITION

COMPANY

ADDRESS

TEL.

LONDON: CHANCERY 5725 BIRMINGHAM: CENTRAL 2626 BRISTOL: BRISTOL 23152 GLASGOW: DOUGLAS 5974 LEEDS: LEEDS 39361 MANCHESTER: DEANS GATE 4638

The two document readers marketed by De La Rue Bull in the U.K. cover a wide range of computer facilities. The LD1, the first of these readers, is used in conjunction with the Bull General Electric Gamma 10, a punched card computer designed for small-to-medium organisations. The MRS-200, however, is a reader/sorter manufactured by General Electric Company (U.S.A.) and which can be used both with their medium-scale Compatibles/400 range and with the large Compatibles/600 systems.

LD1

This reader ($42\frac{1}{4}$ " long, $22\frac{7}{8}$ " wide and 4 ft. high) handles documents printed with the CMC7 characters designed by Machines Bull. CMC7 includes not only a full set of numerals and symbols, but also the complete alphabet. Each CMC7 character is made up of 7 vertical bars separated by 6 spaces. Two space widths are used: long and short. The character code is recognised by the LD1 according to the disposition of the long and short spaces within the characters: the reader registers not the bars but the spaces between them. (This means that printing tolerances are wider with CMC7 than with a code which relies for its recognition on providing the reader with solid characters).

CMC7 characters correspond closely to normal printed characters, and can therefore be easily read by the naked eye. Each numeric character consists of a combination of 2 long and 4 short spaces, giving 15 possible combinations—10 numerals and 5 special symbols. Alphabetic characters, on the other hand, comprise 1 or 3 short spaces and 5 or 3 short spaces: this basic variation gives a full set of 26 letters.

The reading speed of the LD1 varies according to the number of lines of characters on the document: with a single line, 300 documents a minute; with two lines, 150 documents a minute. Various formats can be read, but not in mixed batches.

The normal use of the LD1 reader is for document-to-card conversion.

MRS-200 (Fig. 43)

It is possible to use more than one of these readers on-line simultaneously, though the MRS-200 can also function off-line. The speed of reading is not determined by document size, the MRS-200 handling 1200 documents of mixed sizes a minute.

These reader/sorters operate simultaneously with the central processor and the other peripheral units. The central processor has approximately 50 milliseconds to make a pocket decision, and the MRS-200 sends a signal to the processor 2 milliseconds before each read command. Issuing the new read command before the critical time in this way ensures that the reading speed is maintained at 1200 documents a minute.

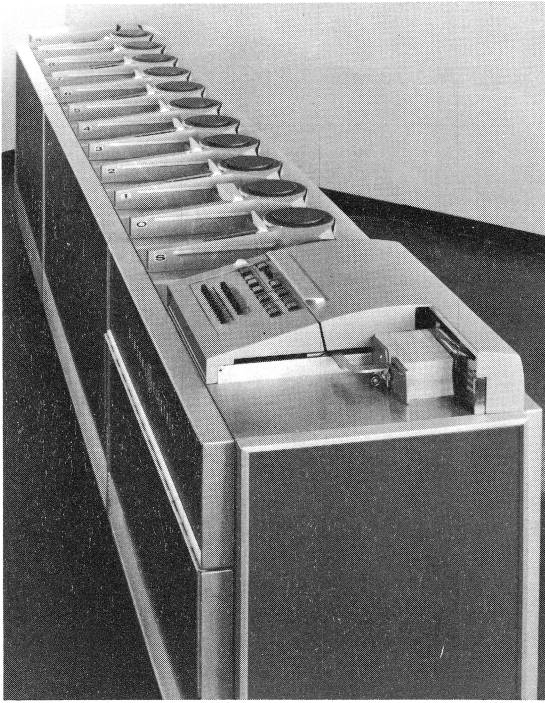


Fig. 43 CMC-7 G.E. MRS-200

In a normal sort, any of 12 digit positions in any of 10 fields may be selected for sorting, and the use of a plugboard enables the reader to handle a wide variety of digit and symbol configurations in any field. (When used off-line the operator uses a control panel to select field and digit). Sort time is reduced by systematic recognition of insignificant zeros and irrelevant data.

There are three main built-in checking features of the MRS-200: symbol check, multiple reading, and missing digit detection. These are used respectively for checking the presence and accurate reading of symbols; for detecting unreadable characters and rejecting the appropriate documents; and for rejecting documents with missing or marginally read characters. Special types of document can be identified and selected by the Multiple Digit Select device.

One important optional feature is the transposition check digit verifier, for checking transposed digits in account numbers.

Used on-line, the MRS-200 can call on ten checking facilities: Transfer timing alert, validity alert, no read command, no pocket command, feeder/pocket alert, document jam, feed failures, missing digit detection, amount field verification, transposition check digit alert. Off-line, there are seven such checks: Symbol check, valid character detection,

missing digit detection, transposition check digit alert (optional),
feeder/pocket alert, document jam, sequence check.

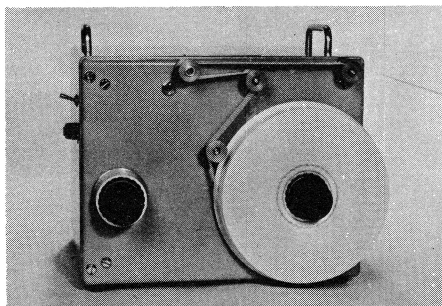
The dimensions of the MRS-200 are 42" high, 28" wide and 176" long.

For further information please contact: -

Mr G. E. Thorne
DE LA RUE BULL MACHINES LTD
114/118 Southampton Row
London WC1

DON'T GET TIED UP WITH TAPE

—Wind it up with a **UNIVERSAL TAPE SPOOLER**



**AS IT COMES OUT OF YOUR
READER/PRINTER/PUNCH**

U.T.S. 150.—UNIVERSAL TAPE SPOOLER
FULLY AUTOMATIC PAPER TAPE WINDER
FOR ALL TYPES OF TAPES AND TALLY
ROLLS—TAKES UP TO 1000FT. LENGTHS AND
WINDS INTO A SELF-SUPPORTING COIL (NO
FLANGES REQUIRED) AND THEN REWINDS
AT HIGH SPEED.

ALSO SPECIAL PURPOSE BUILT CONVEYORS FOR CARDS AND
TAPES TO SPEED THE FLOW OF YOUR DOCUMENTS FROM

DOCUMENT TRANSPORTS LTD

254 HIGH STREET, UXBRIDGE, MIDDLESEX Tel: Uxbridge 37768

English Electric Leo Marconi

Lector

The unique document reader
for use with any computer

Lector, from English Electric Leo Marconi, makes data preparation faster and easier than ever before. Lector reads documents marked by hand, computer printer or address plates; it converts the information into paper tape ready to be fed directly into a computer.

Data can be handled at a rate of over 100,000 characters per hour—over 20 times faster than manual data preparation. An advanced form of this equipment, known as Autolector is now also available for fully automatic, on-line data preparation work.

Applications for Lector include ordering and invoicing, time sheets, payment control, stock control, production control, material handling, piece-work recording and market research.

English Electric-Leo-Marconi Computers Ltd.,
Portland House, Stag Place, London, S.W.1.
Telephone: VICTORIA 2299.



Lector—Automatic Document Reader

In developing the Lector document reader (Fig. 44) English Electric Leo Marconi have provided a highly flexible system for capturing data at source in a form suitable for converting automatically into paper tape input for the computer.



Fig. 44

The equipment comprises a hand-fed document transport, a reading head incorporating 16 photo-electric cells and a paper tape punch.

The documents may be printed on a wide variety of ordinary papers and in pre-determined marking positions marks may be made with a pencil, a computer printer, or an address plate. Marks pre-printed on the form may also be incorporated and the documents can be numbered serially by the printer of the forms both in 'clear' and in bar-coding for reading by the Lector.

The basic document has 16 marking columns across the form and up to 80 lines or rows, down the length of the form, but each document is designed to give the best layout for the system for which it is being designed. In many cases a standard document can be designed to be used in several different applications.

DEALERS NAME AND ADDRESS			ACTION		DEALER		REF. NO.		
140	PRICE	DESCRIPTION	UNIT	QUANTITY DELIVERED				K S D	
8	12/-	JF CFF CVF PR JOLYROLS X 12		1	2	3	4	5	6
10	1/9	WV JF CVF PR CF SWISS ROLL		1	2	3	4	5	6
12	1/9	CVF CFF CADDY ROLLS		1	2	3	4	5	6
14	3/-	R. B. G. O. CHOC ROLLS X 12		1	2	3	4	5	6
16	11d	SPONGE CAKES X 4		1	2	3	4	5	6
18	1/8	SPONGE CAKES X 8		1	2	3	4	5	6
20	1/11	GOLDEN CHOC SPONGE BARS		1	2	3	4	5	6
22	7/6	FRENCH JAM SANDWICH X 3		1	2	3	4	5	6
24	7/6	HAZELNUT SANDWICH X 3		1	2	3	4	5	6
26	6/9	RASPBERRY JAM SANDWICH X 3		1	2	3	4	5	6
28	6/9	COFFEE SANDWICH X 3		1	2	3	4	5	6
30	6/9	LEMON SANDWICH X 3		1	2	3	4	5	6
32	1/9	JF. & VAN. SPONGE SANDWICH		1	2	3	4	5	6
34	1/2	FRUIT FLAN CASES		1	2	3	4	5	6
36	8/8	HARVEST ANGEL X 4		1	2	3	4	5	6
38	8/8	HARVEST MADEIRA X 4		1	2	3	4	5	6
40	11/8	HARVEST SULTANA X 5		1	2	3	4	5	6
42	12/8	HARVEST GENOA X 5		1	2	3	4	5	6
44	12/8	HARVEST DUNDEE X 5		1	2	3	4	5	6
46	1/6	PLAIN MADEIRA		1	2	3	4	5	6
48	1/3	SUGAR'N SPICE GINGER		1	2	3	4	5	6
50	1/11	CH. B. L. D. P.BARGATTO		1	2	3	4	5	6
52	8/8	BATTENBERG CAKE X 4		1	2	3	4	5	6
54	2/6	V/N S. CH.FROSTIE LAYER CAKE		1	2	3	4	5	6
56	8d	A.B./CAA,APR.BAA,SPA,PLA,G.I,ND FR.PIES		1	2	3	4	5	6
58	1/2	CH. L/O. CUP CAKES X 4		1	2	3	4	5	6
60	1/4	TREACLE TARTS X 6		1	2	3	4	5	6
62	1/5	JAM TARTS X 6		1	2	3	4	5	6
64	1/9	ASSORTED TARTS X 6		1	2	3	4	5	6
66	1/9	BAKEWELL TARTS X 6		1	2	3	4	5	6
68	1/9	ICED FANCIES X 6		1	2	3	4	5	6
70	1/11	JERSEY SLICES		1	2	3	4	5	6
72	7/6	CHOC FLAV SANDWICH X 3		1	2	3	4	5	6

20%

L 303/3A. REVISED

Fig. 45: Pre-printed Order Form

TELEPHONE: COUNTY 548

1045243
PLEASE QUOTE THE ABOVE
NUMBER IN ANY COMMUNICATIONS
TO THE BOARD

• • • GREEN • • •

FLAT 6, GILMORE MANSIONS, HILL ROAD,
LONDON, NE1

CLAMSON PARAGON BUSINESS SYSTEMS FORM

FOR KEY TO
TARIFF
CHARGES
SEE OVER

PREVIOUS READING		CURRENT READING	PRIVATE UNITS PER UNIT	SECONDARY OR FLAT RATE UNIT	UNIT CHARGE			WIRE CHARGE			TOTAL CHARGES			
DAY MONTH YEAR					£	s.	d.	£	s.	d.	£	s.	d.	
7	JAN	65	6.5	1000	1230	30	19	6	3	0	0	33	19	6
				SECONDARY FLAT RATE PRIMARY		SECONDARY						TOTAL DUE		
				3.2								33 19 6		

SPECIMEN FORM

ENGLISH ELECTRIC LEO COMPUTERS LTD.

**TOTAL
DUE**

33

198

REGIONAL ELECTRICITY BOARD
MARKET STREET, COUNTYTOWN

R. L. GREEN, JR.

PAYMENT

BY CHECK THIS COUNTERFOIL MUST BE SENT WITH A CROSSED CHECK TO THE ABOVE ADDRESS. NO RECEIPT WILL BE ISSUED UNLESS REQUESTED.

BY BANK CREDIT TRANSFER COMPLETE THE FORM OVERLEAF AND MAKE PAYMENT AT ANY BANK.

BY CASH
THE LIST OF BOARD OFFICES IS SHOWN OVERLEAF. BRING THIS ACCOUNT.

[illegible]

FOR OFFICE USE ONLY

[illegible]

458902243
CONSUMERS NUMBER

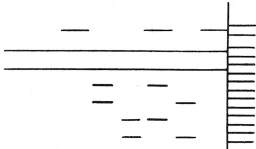
3519	AMOUNT
------	--------

5 JAN 65
DATE

ENGLISH ELECTRIC LEO COMPUTERS LTD.

Fig. 46 Public Utility Bill

Date :
INVOICING COPY



0395

CURRY'S LTD.,
77, UXBRIDGE ROAD,
EALING,
LONDON, W.5.

A.C. No.

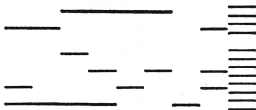
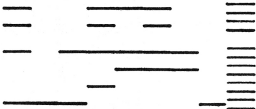
026.292.4

DELIVER TO:-

D

ORDER NUMBER

F	C	W	B	R	V	N
---	---	---	---	---	---	---

QUANTITY	DESCRIPTION	Special Price
1	26x2.00 DUNLOP AUTO CYCLE COVERS 1852416 	1
84	CASINGS W/C 1890001 	2
		3
		4



Sales Checker: D. T. 700

Fig. 47 Delivery note

A plugboard and certain automatic facilities in the Lector allow any number of the 16 columns to be active on a particular document, and pre-printed timing marks controlling the reading of the lines allows complete freedom in spacing the lines along the form.

The documents may be of any size between 5" × 6" and 10" × 18" as long as the length exceeds the width by 1", and documents of different sizes and layouts may be mixed.

Forms are fed by hand into the document transport, and carried under the reading head.

When a timing mark is detected the reading system is switched on and punching of the marks detected on the line begins. Marks may be present in all the marking locations.

During reading, automatic checks are carried out to detect marks of poor quality and to ensure that all marking locations have been examined. A reject bin is provided.

Output is on 5, 7, or 8 hole paper tape at 110 characters per second.

The information marks are translated into holes in the paper tape, 4 columns on the form representing 1 character on the tape. Consequently the bit pattern on the tape provides an indication of where the marks occurred on the form, and the computer program can interpret this pattern according to the layout of the document from which the tape was produced. As the tape is in a non-standard code it can readily be accepted by any computer.

Document feeding rates depend on the size of the document and on the number of characters which have to be punched. A 12" document using 16 columns will be fed at the rate of 1200 documents per hour and give a punching rate of over 100,000 characters per hour.

Over 50 Lectors are being used in this country and overseas and systems already in operation include payroll, ordering, stock control, rent collecting, billing and invoicing applications.

The three typical Lector forms shown indicate the flexibility of form design for use with a Lector system:

Form 1. This is a pre-printed order form used by van salesmen. In this case, only eight columns are needed for the normal range of business work. Easily marked by hand for picking and order recording. Any special requirements are simply written out in full and the 'action' box marked.

Form 2. This public utility bill is printed by the computer, with the payment slip bar-coded for automatic reading on return by the customer. Where the payment is made in full, no additional data entry need be made, part payment details can be marked by hand.

Form 3. This delivery note is produced by address plates and serves immediately both as the customer document and data for the computer.

With an address plate system, complex commercial data can be handled accurately without difficulty; and plates supplemented by bar-coding can provide the complete set of data for the computer automatically.

For further information please contact: -

Mr D. H. Noble
ENGLISH ELECTRIC-LEO-MARCONI COMPUTERS LTD
Portland House
Stag Place
London SW1

COMPUTER FODDER

**How to produce it quickly, accurately and economically
without an army of punch operators.**

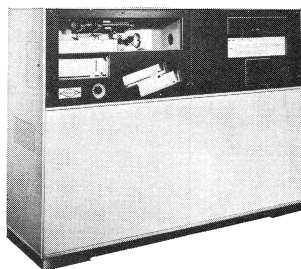
Computers swallow punched tape, magnetic tape or cards quicker than Italians eat spaghetti. If you have to keep a computer busy all day, either you need an army of operators (all of whose output must be checked) or a Farrington Optical Character Reader, (which checks its own work).

The Farrington O.C.R. will accept printed or typed documents (preferably, but not essentially, in its own perfectly readable typeface) and will convert each character, letter or figure, into computer language, using whatever code your computer is accustomed to. It will do this entirely automatically using a highly sophisticated optical reading system at rates of up to 400 characters per second. A series of built-in checking devices continually monitor the machine's performance.

If you have problems of manual conversion we think we can prove to you that a Farrington Optical Character Reader would do you a better job at lower cost.

Ask us to come and justify our claim.

THE FARRINGTON OPTICAL CHARACTER READER



MODEL 3030 PAGE READER

*Please let us have more details of the
Farrington Optical Reader.*

NAME..... TEL NO.....

COMPANY.....

ADDRESS.....

BCS.1.

ADREMA LIMITED

BRADMA HOUSE 65 UXBRIDGE ROAD EALING · LONDON W5 · GEORGIAN 4321.

6. FARRINGTON

Model 3010

General characteristics

Source Documents—reads a single line of numeric data on documents prepared by conventional business machines—high speed printers, typewriters, adding machines, cash registers, credit card imprinters etc.

Multi-Font—accepts Farrington Selfchek 12F and 7B, IBM 1428 and OCR 'A' fonts.

Infra Red—enables machine to ignore visible form lines, overstacking etc. in the reading zone.

Selection of data to be recorded in output.

Right Register—permits right justification of variable length fields.

Self-Checking Fields—the use of check digits to form self-checking fields provides for error free operation.

Mark Guide Data Entry—information may be added to the output record or the program altered by filling a mark guide with pencil or scannable ink.

Set Switch Data Entry—constants may be added to the output record by means of operator-adjustable set switches.

3-Pocket Sorting.

Plugboard Program—easily interchangeable wired plugboards provide programs for each different application.

Counters indicate total documents processed and the contents of each of the output stacking pockets.

Paper and card stock—minimum substance weight 14 lbs—thickness range 0.002 ins. to 0.010 ins.

Document Size Range—2.0 ins. to 6.0 ins. height, 2.25 ins. to 8.5 ins. width.

Output—magnetic tape, punched cards or on-line to computer.

Speed—up to 440 documents per minute.

Optional features

Batch Header—each output record subsequent to a batch header document may include batch header information.

Matrix Mark Scanning—reads data manually marked by pencil or pen in the same pass as that of character recognition.

The Alphabet in capitals may be added to the 12F numeric recognition set.

Two-Line Reading.

Accumulation and Balancing—fixed or variable amounts are accumulated and totals or batch balances transferred to output record.

List Printer—information received from the Reader or Accumulator is printed on a tally roll.

Models 3020 and 3022

General characteristics

Read and Punch—tabulating cards used as source documents/turn around documents, prepared on high speed printers, typewriters, adding machines, cash registers and imprinters are read and the data punched into the same card at a single pass.

Multi-Font—(3022)—accepts Farrington Selfchek 12F and 7B, also IBM 1428E fonts. Single font (3020).

Selective Reading—permits selection of data fields to be read and punched.

Self-Checking Fields—the use of check digits to form self-checking fields provides for error-free operation.

Unreadable Character Fill—provides capacity to determine an unreadable character in a self-checking field and punch it into the card.

Mark Guide Data Entry—information may be added to the documents or the program altered by filling a mark guide with pencil or scannable ink.

Set Switch Data Entry—constants may be punched into each card by means of operator-adjustable set switches.

2-Pocket Sorting.

Plugboard Program—easily interchangeable wired plugboards provide programs for each different application.

Counters indicate the number of documents placed in each of the two output stackers and the total.

Speed—51 and 80 column cards are read and punched at the rate of up to 550 cards per minute.

Post-Punch Checking—each card is checked after punching to ensure accuracy.

Continuous Load and Unload.

Stacker Capacity—input stacker 2000 cards—output stackers 1500 to 2000 cards.

Optional features

Batch Header—each output record subsequent to a batch header document may include batch header information.

Matrix Mark Scanning—reads data manually marked by pencil or pen in the same pass as that of character recognition.

The Alphabet in capitals may be added to the 12F numeric recognition set.

Accumulation and Balancing—fixed or variable amounts are accumulated and totals or batch balances punched into a batch trailer card.

List Printer—information received from the Reader or Accumulator is printed on a tally roll.

Batch Numbering—provides for automatic punching of a batch number into each card in a batch.

Serial Numbering—provides for automatic punching of a serial number into each card passing through.

Model 3030

General characteristics

Off-Line General Purpose Input Data Conversion—the Page Reader automates computer input independently of computer main frame time and programming expense.

Alpha/Numeric Recognition—ten numeric, twenty six alphabetic characters, seventeen signs and punctuation together with six control symbols.

Multi-Line Reading—up to 70 lines per page.

Infra Red—enables the machine to ignore visible form lines, over-stamping etc. in the reading zone.

Selection of data to be recorded in output by line or data field.

Variable Length Line Sweep—saves reading time and increases throughput.

Skip-to-mark—only lines containing data to be recorded in output are read. Other lines are skipped over at high speed.

Mark Guide Data Entry—information may be added to the output record by filling a mark guide with pencil or scannable ink.

Set Switch Data Entry—constants may be added to the output record by means of operator-adjustable set switches.

2-Pocket Sorting.

Plugboard Program—easily interchangeable wired plugboards provide programs for each different application.

Counters—indicate number of lines, records, pages etc. passed through.

Document Size Range—4.5 ins. to 8.5 ins. width, 5.62 ins. to 13.5 ins. length.

Output—magnetic tape, punched paper tape, punched cards or on-line to computer.

Reading speed—40 line-inches per second

Optional features

Reject Insertion—characters rejected by the machine may be keyed in by the operator—output records are complete.

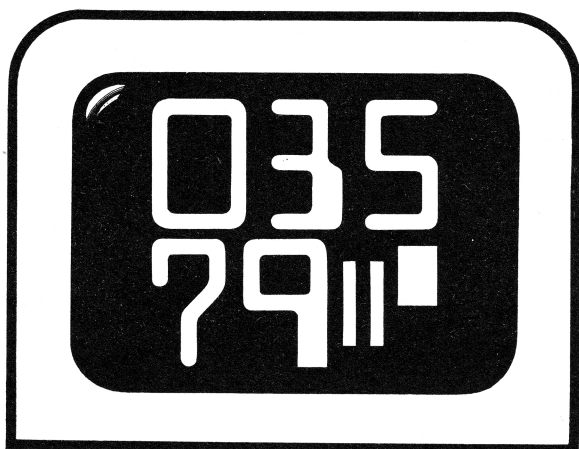
Right Register—permits right justification of variable length fields.

Underscore—provides the capability to encode upper and lower case characters in the output.

For further information please contact: -

Adrema Ltd OCR Division
Bradma House
65 Uxbridge Road
Ealing W5

Tel No: GEOrgian 4321



Some character references for Honeywell M.I.C.R.

Honeywell's Series 200 computers benefit from a comprehensive range of standard peripheral equipment, including an advanced M.I.C.R. Sorter-Reader — widely used and proven where optimum character recognition input is required — and a variety of video display systems.

The capabilities:

- Reading approximately 1600 documents per minute
- Thirteen pocket sorting (ten digital, two special items; and one reject pocket)
- Document loading and unloading during full speed sorting
- Simultaneous operation with high-speed printers or multiple tape lists,

give Honeywell M.I.C.R. Techniques wide application possibilities.

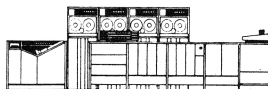
To assist in applying the Series 200 to particular needs, Honeywell has designed, developed and perfected a

powerful range of data handling aids, including:

- The M.I.C.R. Entry package, part of Honeywell's Demand Deposit Accounting System, embodies advanced concepts based on years of experience in leading banking institutions.
- Honeywell VIP System (Visual Information Projection), a versatile solution to the problem of quick retrieval of pertinent information in a usable and manageable form. VIP offers three types of visual display station to accommodate a wide variety of applications and data handling needs.

The Honeywell people have engineered a concept in Data Processing which embodies not only modularity in systems, but also ensures full benefits from available software and peripheral equipment throughout the Series 200 range.

Honeywell Controls Ltd., Honeywell House, Gt. West Rd., Brentford, Middx.



Honeywell

IBM Document Handling and Character Recognition Equipment

The cost of data conversion from source documents into forms suitable for data processing can, on occasion, be as much as the actual cost of processing it. Recognising this, IBM has designed a number of document handling and character recognition devices; the ability of such devices to read printed or marked documents automatically for computer input has reduced the time and cost of data preparation considerably, and by so doing has provided many users with an extended range of practical data processing applications.

Experimental devices have been demonstrated which are capable of reading characters printed in a number of styles and handwritten characters, but no machine has yet been developed that is suitable for general application. These pages describe some of the most important IBM machines currently available (but see p. 122).

IBM Magnetic Ink Character Readers

The number of cheques written every day continues to increase. One bank alone has to clear half a million every working day. To cope with this volume and thus provide better customer service, banks in this country have adopted the magnetic ink character recognition technique; the basis of this system is a line of magnetic ink characters printed in E 13 B (some overseas banks use CMC 7) across the bottom of each cheque. IBM machines under stored-program control read these characters, and cheques can be sorted automatically and the data transferred directly into a computer. Documents are handled faster, labour is saved, and the data can be processed more quickly.

IBM 1419 Magnetic Character Reader (Fig. 48)

This machine, an advance on the IBM 1219, combines fast operation with a high degree of accuracy. All characters in all fields are checked for validity—in fact such is the accuracy of the unique IBM multi-channel method of character sensing that positive identification is possible even though a portion of the character may be missing. The 1419 can read up to 1,600 documents a minute; sort and select documents bearing a given account number or several given transaction numbers; edit documents to ensure that they contain a given transit number; accumulate and print totals for batch proving; and check documents for correct sequence. It operates under stored program control with a minimum processing time (including pocket selection) of 32.2 milliseconds; the 1419 can also be used off-line as a document sorter. There are three models available.

The IBM 1412 Magnetic Character Reader provides similar facilities, though it operates at slightly slower reading and sorting speeds.

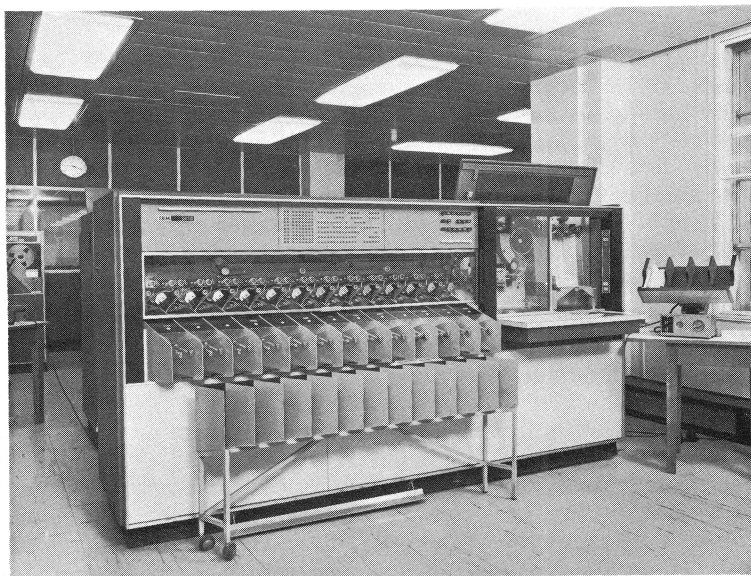


Fig. 48 IBM 1419 Magnetic Character Reader installed at the National Provincial Bank.



Fig. 49 IBM 1231/1232 Optical Mark Reader

IBM 1445 Printer

The chief application of the IBM 1445 on-line printer has been encoding cheques and other documents with magnetic ink characters. It provides extremely high-quality E 13 B or CMC 7 magnetic ink character—or conventional—printed output on continuous stationery. Actual printing speeds will depend on the work in hand, from a rated 190 lines a minute using the standard 56 alphameric character set (including 14 symbols in CMC 7) to 525 lines a minute with a 14 numeric character set. The computer stored program performs all format and analysis control, and a 12-channel carriage tape controls spacing and skipping.

IBM Optical Mark Reading Equipment

There are many applications where magnetic ink printing is not practicable, and for these IBM makes optical character and optical mark readers, which read information printed on documents or pencil-marked on code sheets directly into a computer. This eliminates the need to key punch data from such documents into cards for computer input. The field of application is a wide one—order entry, insurance premiums, stock control, market surveys, and payroll, to list just a few.

IBM 1231 and 1232 Optical Mark Readers (Fig. 49)

Both the IBM 1231 and the IBM 1232 Optical Mark Readers read pencil marks from a source document, the 1231 transferring this information directly into a computer, the 1232 converting it into punched cards. The data for processing therefore only needs to be recorded once at source, by personnel who do not need extensive training.

Source sheets are $8\frac{1}{2}$ by 11 inches, with up to 1,000 mark positions grouped into 100 words of 10 characters each. They can be marked with an ordinary lead pencil (2B is recommended) or by an IBM 1403 on-line printer, and fed automatically from the hopper. The Mark Discrimination feature will automatically reject sheets with weak or invalid characters. Programming is simple: specially marked sheets are loaded into the machine ahead of the data sheets, and a sonic delay line stores program data until it is transferred to the computer or card punch.

IBM 1231 This machine reads, checks, and transfers information directly into a computer at up to 2 million characters per hour. A switch setting determines continuous (2,000 documents per hour) or demand feeding (under computer control at up to 1,600 per hour). The machine reading operation is fully buffered.

IBM 1232 This machine transfers data at up to 1,400 documents an hour via an IBM 534 Card Punch into 80-column cards—the equivalent of 68,000 keypunch key strokes an hour. Common information can be duplicated by the 534 into a group of cards; when not connected with the 1232 the 534 can be used independently. Certain optional special features provide additional operating versatility.

IBM 1285 Tally Roll Reader

The 1285 will benefit organisations such as wholesalers, discount houses, local government authorities, hire purchase companies, building societies, etc.—wherever there is a need for a regular fast, accurate reporting procedure to management of day-to-day cash counter trans-

actions. The 1285 reads data for processing directly from cash register tally rolls at a maximum rate of 3,000 lines a minute into an IBM computer. It uses a unique electronic flying-spot scanner consisting of a cathode ray tube, optics, and photomultipliers which will read from right to left up to 25 characters on one of a number of lines, without moving the roll being scanned. Line positioning is automatic, the operator controlling the fields to be scanned by setting the display screen dial.

The 1285 accepts characters in IBM 1428 numeric font plus C, N, S, T, X, and Z or (with the addition of a special feature) the numeric optical recognition font developed by NCR. Unreadable characters are automatically rescanned and can be corrected on-line by the operator or marked automatically for off-line correction. Automatic vacuum loading and tightrewind of the tally roll ensure correct positioning, prevention of tearing, and the easy renewal of the roll. The 1285 is controlled by the processing unit of the computer system, and up to eight 1285s can be attached to an IBM System/360.

IBM 1428 Alphameric Optical Reader

The IBM 1428 has been designed to facilitate the handling of large volumes of documents and routine paper work. Typical applications include meter reading transcription, insurance premium payments, instalment loan accounting, subscription renewal, tax accounting, and dividend warrants. The 1428 reads alphabetic and numeric data printed by the IBM 1403 Printer or by an IBM Electric Typewriter with 1428 type font. Accuracy of character recognition is ensured by the 1428's high resolution scanning system, and since the 1428 operates under stored-program control operator intervention can be reduced to a minimum.

The 1428 recognises 42 characters (upper-case A to Z, the digits 0 to 9, and the more commonly used special characters) at up to 480 characters per second. Documents are read at up to 400 per minute; they should be between $2\frac{1}{2}$ and 3 inches high and between $4\frac{7}{32}$ and $8\frac{3}{4}$ inches wide.

There are three models available. Models 1 and 3 have 3 output stackers; Model 2 has 13, and is thus suitable for off-line sorting of paper documents. Optional special features include a mark reading station for reading information coded in pen or pencil as vertical strokes, or an additional read station which allows information to be read from any two lines (standard and additional) on a document in a single pass through the machine.

IBM 1418 Optical Character Reader (Fig. 50)

The 1418 differs from the 1428 in that it can read only numeric and certain special characters. Its speed is the same, 480 characters a second directly into a computer; it can also be used off-line for fast sorting. As with the 1428, reading selection, zero insertion and field checking are all under computer stored-program control.

IBM System/360

An important feature of the IBM document handling and character recognition equipment described above is that most of them can be at-

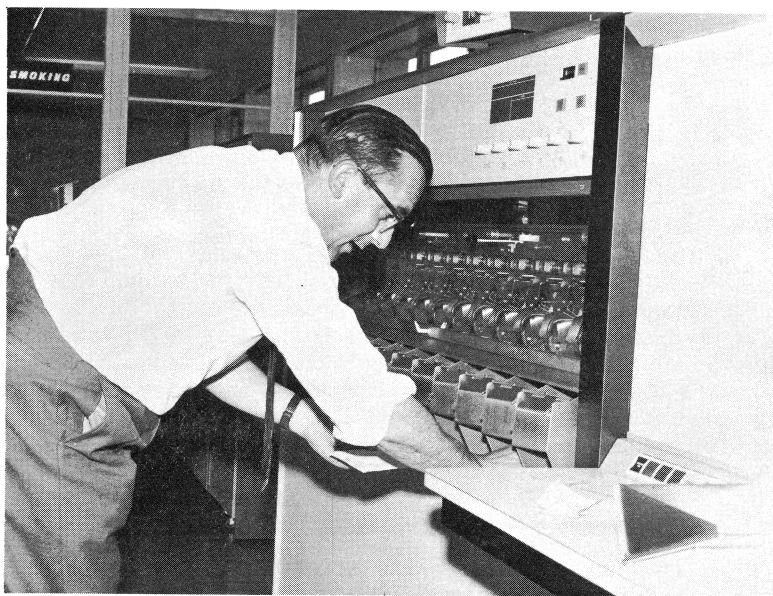


Fig. 50 IBM 1418 Optical Character Reader installed at the East Midlands Gas Board.

tached by the standard input interfaces to IBM System/360. Operating System/360 which provides Control, Processing and Service Programs and flexible input/output channel facilities enable the computer to overlap processing the peripheral operations with the transfer of data into the system. The control units provide the logical capability to operate and control the I/O device and to adapt each device to the standard form of control provided by the selector channel being used. Thus concern with I/O device assignment is minimised, and document handling and character recognition devices can be incorporated into the system and programmed with the same easy facility as any other peripheral device. This integration has led to the creation of a powerful management control tool, and is already having a profound effect on the development of new management techniques.

For further information please contact:-

Mr M. Belman
I.B.M. United Kingdom Ltd
40 Basinghall Street
London E.C.2.
Tel. NAT 8051

This article was written before the announcement of the IBM 1287 Optical Reader, which reads handwritten numerals as well as numerals printed in several styles. Information is available from the address above.

In a commercial environment the amount of data handling is large compared to the amount of computation. As computers have improved in their capability to store, retrieve and handle large quantities of data, the ever present problem of data preparation has grown. Where large volumes of data have to be prepared, speed and reliability are of paramount importance—and each hole punched in a card or paper tape introduces a possibility of human error.

To alleviate this obstacle, I.C.T.'s systems designers have included, in the 1900 Series of computers, a range of original document handling devices.

Universal Document Transport (Fig. 51)

The I.C.T. Universal Document Transport is a flexible computer input device for data held on a wide variety of paper sizes. The data may be in the form of marks in defined areas of the paper, holes in punched cards or, in future variants, alpha-numeric characters of the ISO B font. Several Universal Document Transports may be connected on-line to any of the 1900 Series central processors by means of the I.C.T. Stand-

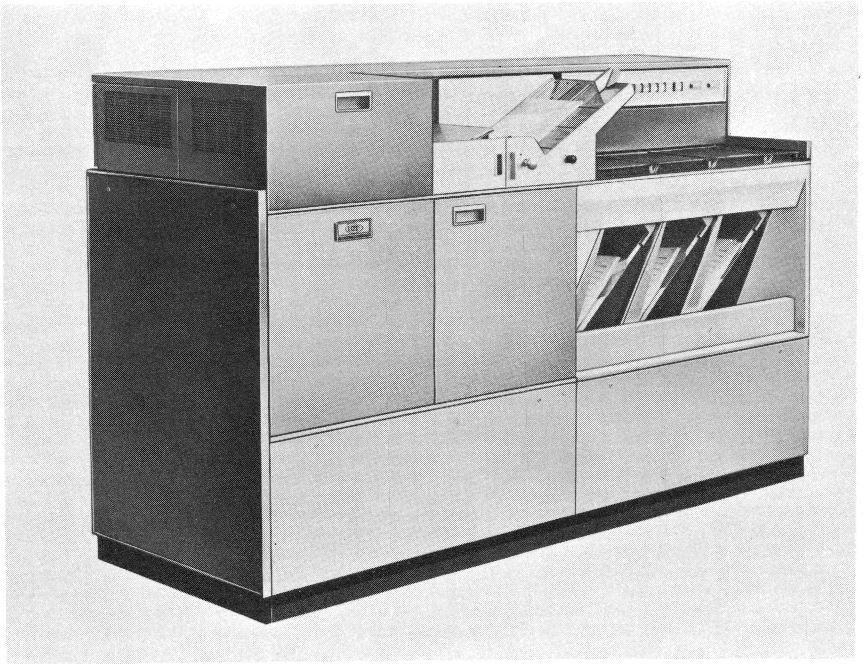


Fig. 51 I.C.T. 1900 Series Universal Document Transport

ard Interface; in such installations they are wholly compatible with multi-program working because of the undemanding nature of their operation.

Capacity and speeds. The Universal Document Transport can handle any document that falls within two size ranges. Small documents are classified as those that are between 5.5 inches long by 3 inches wide and 8.5 inches long by 6 inches wide. Documents of greater dimensions—up to a maximum of 13 inches long by 8.5 inches wide—are classified as large. Small and large documents are read at rates of 300 to 150 documents a minute respectively.

Before a reading sequence is initiated, the size range of the documents to be handled and the form in which the data are recorded are selected. Once this has been done and the U.D.T. started, documents are automatically fed past the reading heads and collected in the computer specified output stackers of the machine. Up to 12 output stackers can be specified in groups of 3.

While initially available with a mark reader, the U.D.T. is designed to accept other readers and an optical character line reader (ISO B font) is already being developed.

Transport. The documents are stacked flat, printed side upper-most, in the feed hopper so that the leading edge of the lowest document rests on a revolving feed drum. At the appropriate instant in the machine operating cycle, a vacuum is applied to ports in this drum and the lowest document in the hopper is moved forward. The document is transported along a tracking unit which aligns the reference edge of the document against an index plate; at the same time sensing devices record the document's passage and confirm that only one document has been fed and that its movement is correctly synchronised with the machine cycle. If an error is detected at this stage, a gate at the end of the tracking unit diverts the faulty document on to a reject platform; this system ensures that machine errors are corrected at the earliest opportunity.

Once accepted into the machine, documents are directed, one at a time, from the tracking unit on to the periphery of a drum, a vacuum within the drum causing them to adhere to its surface. The dimensions of this drum permit four short or two long documents to be held simultaneously.

As the drum revolves, it carries the documents past the reading stages where both the mark and the future optical character line readers may be fitted. Each reader is connected to the computer via a separate channel to enable simultaneous reading of two different data 'fields' on the same document. A document can be retained on the drum for several revolutions so that on-demand reading can take place; when documents are retained, the document sequence remains undisturbed.

This re-circulation offers the flexibility necessary for operation in a multi-programming computer configuration. It permits peripheral transfers associated with a particular program to be suspended in favour of those associated with a higher priority program, or document reading to be matched to the performance of particular direct-access storage devices—so allowing the results of accessing particular storage areas to determine the stacker destination of the document.

When a document is to be released, guide fingers lift it from the drum and divert the document on to a track leading to the output stackers each of which incorporates a computer controlled gate.

Mark reader. This reader can identify marks made by hand, by computer line printer or other means. As many as 23 data columns plus a 'clock' column can be disposed across an 8.5 inch wide document, and up to 60 rows of marks can be accommodated on a document 13 inches in length.

When required, the mark reader can be set to identify the holes perforated in existing punched cards; in effect the reader detects the non-reflecting black surface of the drum visible through each rectangular punching.

When reading a mark, the reader transmits one of three signals to the computer via the Standard Interface. If a clearly defined mark is present, the reader transmits a 'mark' signal; if no mark is present, it transmits a 'no mark' signal; if the reader cannot clearly detect either of these conditions, it transmits an 'uncertain mark' signal. This system permits the computer program to assess the probability of a mark and to act accordingly—in some applications the data format permits an uncertain mark to be rejected or accepted by program of that format. This unique feature enables the peripheral to use the decision making power of the computer.

For the future—an optical character line reader. A single line of standard ISO B font alpha-numeric characters will be read by this reader. This line may be positioned anywhere on the document, the reader being adjusted to scan the appropriate area of the documents prior to the particular run. Documents prepared by computer printer, suitable typewriters and embossed address plates will be identified—the characters will be pitched at either eight or ten to the inch.

U.D.T. as a sorter. Although document sorting is not a primary function of the U.D.T., this device can perform this duty with great efficiency.

8.2 Optical Print Quality Monitor (Fig. 52)

Designed to check that printed characters are of sufficient quality to be read by machines, the Optical Print Quality Monitor can measure paper whiteness and blemishes, and the dimensions, density and quality of print characters against parameters set up beforehand.

A scanning cathode-ray tube is used as the light source, its controlled output being focussed by a lens on to a window in the document platform. This permits one character to be scanned at a time.

The light reflected from the sample character is detected by a photomultiplier which converts it into a video frequency wave form. This is then compared with the reference level set by the operator and the resultant signal is used to re-form the character as a magnified image of the sample. This image appears as extreme light and dark areas on a second cathode-ray tube used as the control panel screen. The properties of the characters being examined are checked against a measured



Fig. 52 I.C.T. Optical Print Quality Monitor.

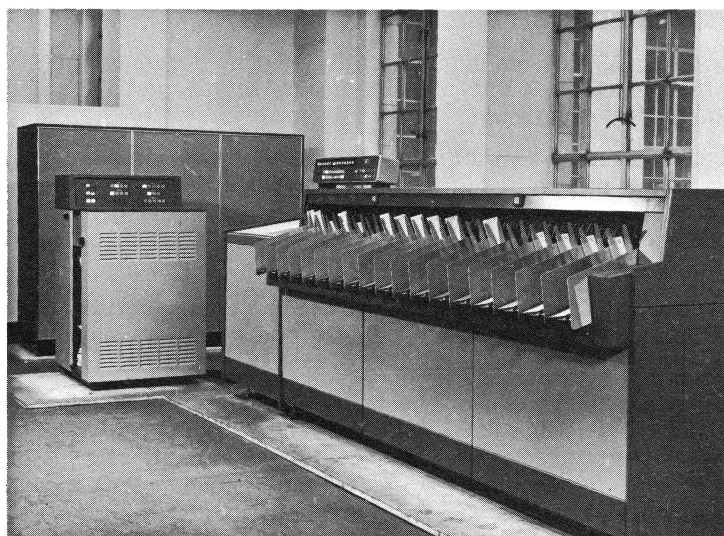


Fig. 53 Document sorter at the National Provincial Bank.

grid or directly compared with the specified character outline engraved on a removable perspex graticule. One graticule is required for each character shape to be checked.

The Optical Print Quality Monitor may be used for simple 'pass or fail' character checking or for the absolute measurement of print parameters (using a graticule calibrated in thousandths of an inch or hundredths of milli-metres).

Document Sorter/Reader (Fig. 53)

The 8500 reads and sorts original documents, of varying sizes and thicknesses at a speed of 1, 200 a minute. Data can be recorded in magnetic ink characters of the E 13 B or CMC 7 fonts. Where subsequent distribution demands the physical sorting of the documents, as apart from the recorded data, the 8500 acts in an off-line role for grouping them into numerical and other sequences at the same speed.

For off-line sorting processes the 8500 can be specified with 6, 12 or 18 stackers. Simple sorting is carried out in digital sequence—units, tens, hundreds and so on—overall sort times being reduced by 'zero kill' and 'high activity selection' techniques on the 18-stacker version. The 8500 can be continuously run, as the feeding of the variously sized documents and the emptying of receiving stackers can be carried out while the machine is in operation.

Through the I.C.T. Standard Interface facility one or more Sorter/Readers can be connected on-line to any central processor in the 1900 Series. Document data is fed directly to the central processor and information can be returned to the 8500 to determine selection or grouping of the documents in the stackers.

M.I.C.R. Adding Amount Encoder 8550 (Fig. 54)

The 8550 feeds, positions automatically, encodes documents in M.I.C.R. characters, produces a tally list of amounts with a total and automatically stacks the documents.

The 8550 encodes an amount field of up to 11 digits, plus the opening and closing field symbols. A full keyboard provides the means of selecting digits 1 to 9 in each of the eleven digit positions. For sterling machines keys are provided for 10d and 11d. Zeroes are encoded automatically in all non-selected digit positions.

As an optional feature the facility of encoding a transaction code may be provided, the transaction code appearing in a specified position. There is also a facility for repeat printing of transaction code.

Single documents are manually fed into the document transport, where revolving brushes align the document to the encoding position. If a document is not correctly registered, encoding is automatically inhibited. After encoding, the document is ejected into a stacker with a capacity of 500 documents. The 8550 may be specified with either E-13B or CMC7 type fonts.

Documents within the following size ranges may be encoded:

Length 6 inches to 8.75 inches

Depth 2.75 inches to 4.5 inches.

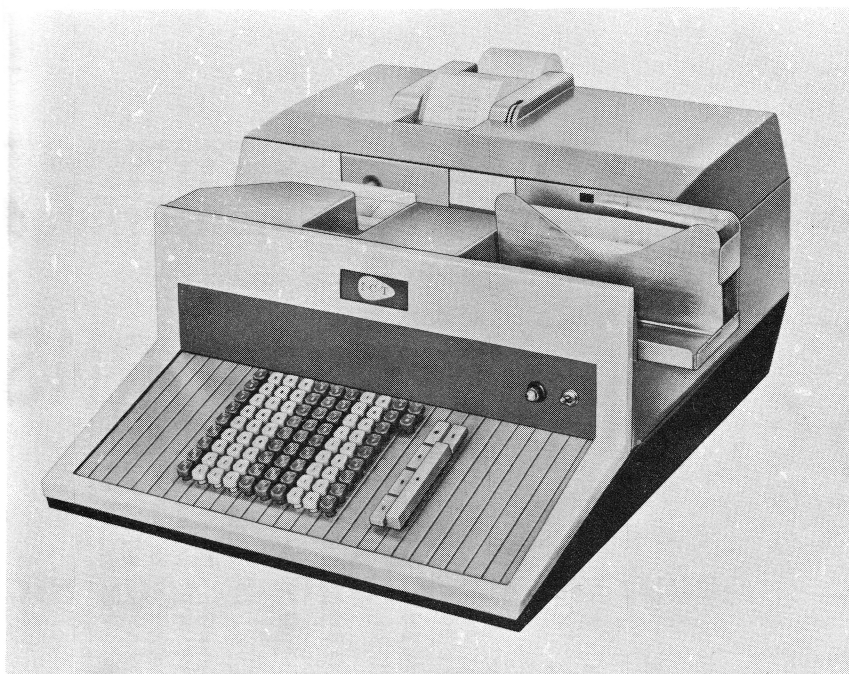


Fig. 54 Adding Amount (check) Encoder. 8550 (Data Recording and capture)

Thickness of documents may be between .003 inches and .010 inches. The M.I.C.R. ribbon has the capacity for encoding approximately 10,000 documents (three day's work). The 8550 has been designed to ensure that ribbon changing may be effected quickly and easily.

For further information please contact:-

Mr. G. R. Bunting
Press Officer
L.C.T. House
Putney
London, SW15.

				ONLY COMPUTER ESTIMATE	H	9	9	9	9	9	9	9	9	9	9
--	--	--	--	------------------------------	---	---	---	---	---	---	---	---	---	---	---

For High-speed - Precision Printed

YEAR	TARIFF	METER NO.	LOCATION	TYPE	CARD	B.P.	SERIAL
1ST QTR.		2ND QTR.					
3RD QTR.		4TH QTR.					

OCR BUSINESS FORMS

Accurate to within $\frac{3}{1000}$ inch

READING	A	2	2	2	2	2	2	2	2	2	2	2	2	2	2
CHECKED	B	3	3	3	3	3	3	3	3	3	3	3	3	3	3
DIAL	C	4	4	4	4	4	4	4	4	4	4	4	4	4	4
CARD	D	5	5	5	5	5	5	5	5	5	5	5	5	5	5
METER	E	6	6	6	6	6	6	6	6	6	6	6	6	6	6
UNGED	F	7	7	7	7	7	7	7	7	7	7	7	7	7	7
METER	G	8	8	8	8	8	8	8	8	8	8	8	8	8	8
DISCTED	H	9	9	9	9	9	9	9	9	9	9	9	9	9	9
CONSUMER															
PREMISES															
VACANT															
RENTS															
ONLY															
COMPUTER															
ESTIMATE															

Call in Lamson Paragon Limited
Britain's Leading Manufacturers of Business Forms

LAMSON PARAGON LIMITED BUSINESS FORMS

PARAGON WORKS LONDON E16

TELEPHONE: ALBERT DOCK 3232

MEMBER OF THE LAMSON INDUSTRIES GROUP

Direct sales and service everywhere

Regional Offices:

BELFAST
BIRMINGHAM
BRISTOL
CARDIFF
DUBLIN
EDINBURGH
GLASGOW
LEEDS
LIVERPOOL
MANCHESTER
NEWCASTLE-ON-TYNE

25004
Midland 4372/3
293321/2
37250
62191
Waverley 6444
City 3438/9
31811/2
Royal 1909
Central 2466
21610/23322

0002340030500 066 -||
 0002340270015 666 -||
 0002340560006 766 -||
 0002340040707 066 -||

WE'RE ON SPEAKING TERMS
 WITH EVERY COMPUTER

||004999|| 78|| 24321:
 13 24968 1|| ||005000||
 78|| 24321: 13 24968 1||

NCR equipment speaks every type of computer language—punched tape or cards, magnetic characters or optical characters.

A computer is only as good as the information it receives. So to make sure it works at peak efficiency it must be fed with completely accurate information prepared at the lowest possible cost.

NCR achieves this through its wide and unique range of input preparation equipment. Facts recorded on NCR adding machines, accounting machines or cash registers are simultaneously converted to paper tape (or other forms of input) in one fast, automatic, self-checking operation.

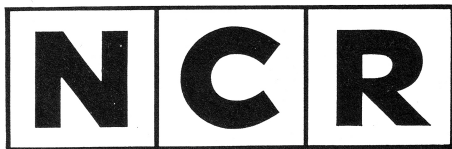
This breaks the language barrier between recording machine and computer. Automation starts the moment the first essential printed record is made—the benefits of electronic data processing are no longer eroded by errors and high hidden costs.

And because NCR input preparation requires neither special skill nor extra effort it gives even small businesses a 'direct line' to a computing centre. Thousands of firms throughout the world are now proving the remarkable speed and economy of 'input by NCR'.

NCR speaks the language of computers. You should speak to NCR—today.

NCR have a unique range of business equipment that can keep you ahead of your competitors—Adding and Accounting Machines, Cash Registers, Input Preparation Equipment, Electronic Computers. . . And in planning your firm's future growth you need NCR equipment and NCR 'knowhow'. **Start that planning now. Talk to NCR.**

plan ahead with



THE NATIONAL CASH REGISTER COMPANY LTD, 206 Marylebone Rd., London NW1, PADdington 7070

NCRMC

As soon as computers began to make their impact on the world of business, NCR realised that the success of these operations would always depend to a large extent on the way in which data was collected at the points of origin in a form suitable for automatic processing.

NCR Character Recognition systems are part of the wide range of data capturing and data preparation equipment that has been developed as a result of that early realisation. They have been designed to operate economically under specific conditions and are often supported by other types of input machine. It is appropriate, therefore, to consider their characteristics in relation to the objectives of the Company's 'total input' programme which are listed below.

1. Data should be recorded in computer 'language' at or near the point of origin, thereby reducing the number of manual operations required, and virtually eliminating transcription errors.
2. Data should be 'captured' wherever possible, as an automatic by-product of other essential recording operations.
3. The recording of data in computer 'language' should be coupled with the production of 'hard copy' records.
4. The validity of data should be fully verified before it is fed to the computer as errors can be more easily corrected at the point of origin.

NCR's development work has for many years been concerned with improving established forms of input preparation; particularly the production of punched paper tape and punched cards. Particular attention has been devoted to data capturing systems which create tape or cards as by-products of normal accounting operations whilst producing visible records for immediate use. In these systems the validity of the captured data is checked by normal accounting methods such as a numerical proof total.

Input preparation units, another important sphere of NCR's activities, change the emphasis of the operation. Hard copy records are produced as a by-product of the data preparation operation whilst maintaining the ease with which input media are prepared. Also, the creation of tape or cards can be the by-product of a conventional listing operation. An important feature of input preparation units is the availability of check digit verification for account numbers and other important reference numbers. The input preparation units provide a full-amount keyboard with punching format controlled automatically. Many thousands of NCR input preparation units are in use throughout the world.

MICR (Magnetic Ink Character Recognition)

The pioneering work in the field of MICR was done in the late fifties by NCR in collaboration with the Bank of America. The result of this

pioneering work was the production of the NCR 402 sorter capable of handling E13 B encoded cheques and other documents at the rate of 750 per minute. NCR's new sorter, the 407, operates at up to double the speed of the 402, giving a handling capability of up to 1,500 documents per minute.

Further development included magnetic amount encoders operated by conventional adding machines of which the latest model, the British built 480-2 (see Figure 55) was built to specifications laid down by the British clearing banks.



Fig. 55 The NCR 480-2 MICR Document Encoder.

Initial orders for encoders worth millions of pounds have already been received from many of the major British banks.

OCR (Optical Character Recognition) and Retail Automation

In common with most computer manufacturers NCR foresaw that the advent of faster and more sophisticated computers created the necessity to be able to input source documents direct to the computer.

The retail business in particular had a need for low-cost point of sale input preparation equipment because of the relatively large number of data originating points.

An obviously readily available source document present in most retail business is the tally roll of a cash register or adding machine.

Accordingly NCR began work some years ago to produce an optical reader capable of reading tally rolls. Also, it was soon realised that as NCR has many thousands of cash registers and adding machines in the field it was necessary that tally roll printing mechanisms be able to be embodied at minimum cost. The requirements were fulfilled by

the provision of mechanisms able to print in a stylised format called NOF (National Optical Font), the incorporation of such a mechanism adding hardly more than £80 to the cost of the basic machine. Figure 56 below shows the NOF 16 character set consisting of the numerals 0 to 9 and six identification and special characters.

The requirements for the optical reader were soon established as:-

1. Read tally rolls produced by existing machinery.
2. Operate at minimum reject rate with maximum checking.
3. Edit information.

In addition, the foregoing three requirements had all to be met at the lowest possible cost.

Bearing in mind the various problems created by the different printing mechanisms with variations in height, width and print quality and alignment it was necessary to find the code most tolerant of the various differences. This was established as being a bi-coded system in which the top and bottom of the character could be decoded. The code itself is so arranged that in order to transform one character into another at least two binary bits must be converted. Statistical studies and experiments have shown that the probability of such an occurrence is extremely small.

Skewing the character to the right or left does not convert one character to another until at least 7.5 degrees is reached. The printing mechanism and paper transport system makes this degree of skew impossible.

Shown below in Figure 57 is the fundamental bi-code character grid structure showing how a vertical stroke in any section of the grid is a binary 'one'.

In the transition from the research to the engineering phase the basic reading concepts were retained. Two apertures would decode the top and bottom zones of the characters. Two more apertures were added, one above and one below the decoding pair, to enable self clocking and reading to be carried out.

A vacuum take up spool to allow rapid loading and unloading of tape was incorporated and an automatic re-read facility was added to minimize the incidence of manual correction required by the detection of a doubtful or unclear character.

The result of NCR's research in this field was the NCR 420-1 Optical Reader capable of reading 1560 lines of tally roll per minute. NCR's newest optical reader, the 420-2 (see Figure 58) has been modified to render it capable of reading up to 52 lines per second or 3120 lines per minute.

Already over 12,000 cash registers and adding machines in the U.S.A. are operating with NOF and over 200 NCR 420 readers are operating to process the rolls. Figure 59 shows a typical section of an NOF tally roll.

In Britain a number of large retail organisations have placed orders for similar equipment and optical character readers are already offering processing facilities in NCR's computing centre in Moorgate, London.

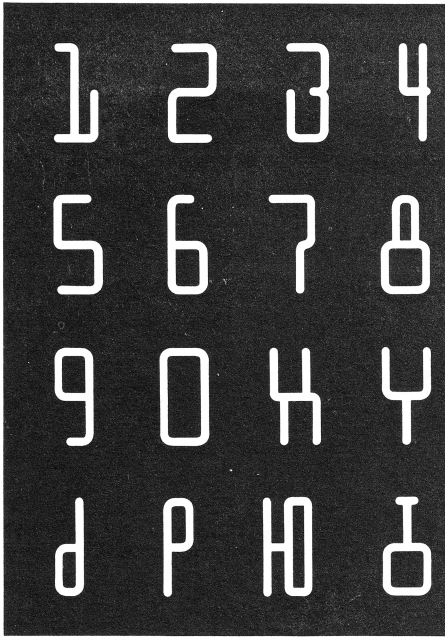


Fig. 56 The NOF 16 Character Set.

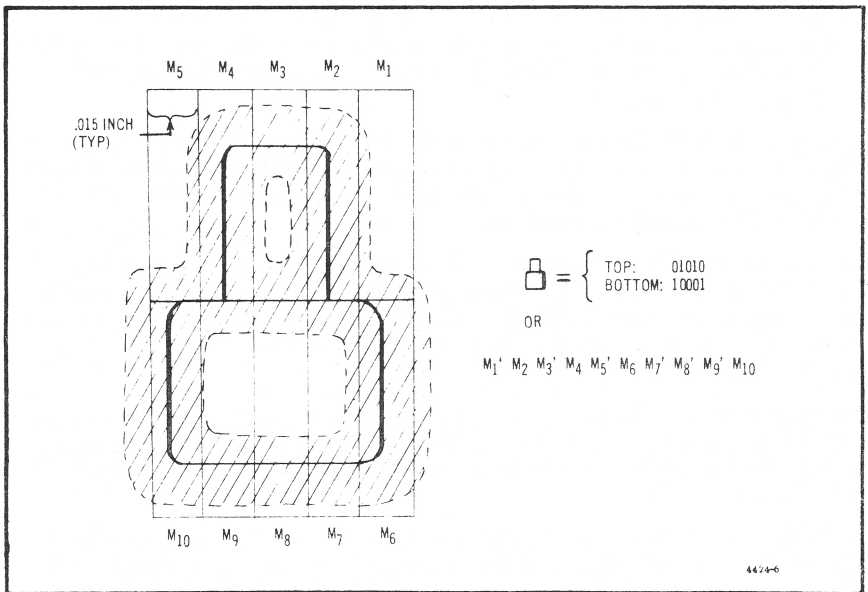


Fig. 57. The bi-code character grid structure

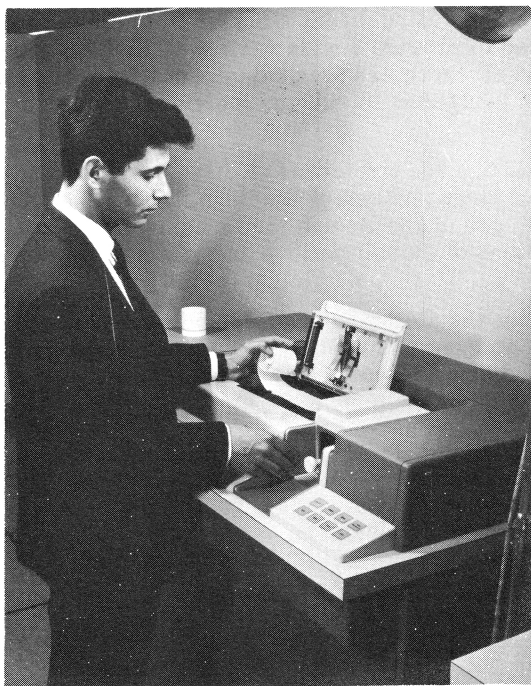


Fig. 58 Loading the NCR 420 Optical Reader

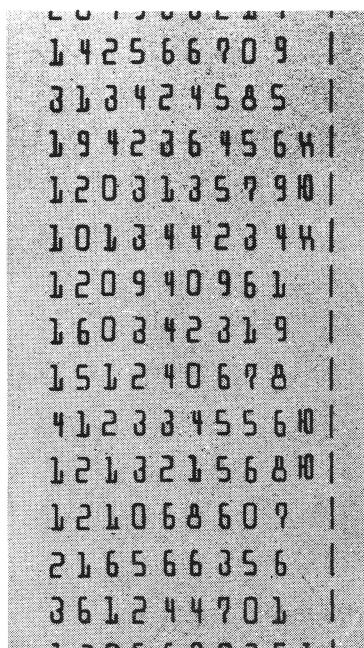


Fig. 59 A typical section of a NOF Tally Roll.

In addition NCR's computing centre offers a series of package programs specially designed to provide a wide variety of NOF tally roll processing facilities. In all there are four basic packages but each one may have several slight variations to suit the individual requirements of the customer.

These package programs provide the means of dissecting and analysing sales information over a large number of headings. These include goods classification with grand and/or sub totals, assistant identification and many combinations of classifications and totals at predetermined intervals.

This information provides the ability to produce analyses under such headings as department and assistant totals, branch totals, goods classification, and detailed breakdowns of total turnover.

Special programs may, of course, be written to meet any customer's individual requirements.

For further information please contact:-

W. G. Sommer
NATIONAL CASH REGISTER COMPANY LTD
206/216 Marylebone Road
London NW1

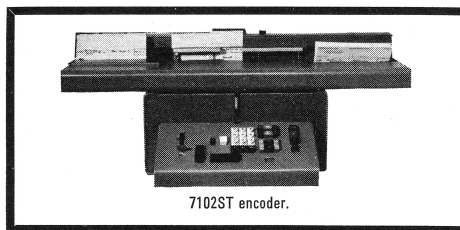
Print in either MICR or OCR without mistakes with Olivetti

Multi-feed encoders with optional automatic feed

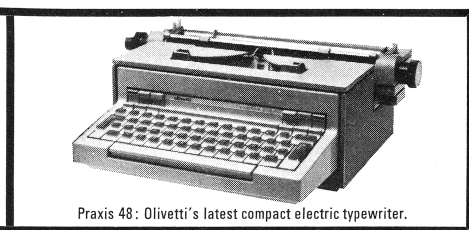
British Olivetti's full range of encoders – sterling or decimal; magnetic character E13B or CMC7; OCR any font; with or without arithmetic – give automatic protection against machine or operator error. Example, the main motor bar locks to prevent encoding when no document is in the printing position; or when the document is in transit; or when field capacity is exceeded; or when the ribbon has reached its 90,000 character capacity. Another example – the automatic elimination of out-of-sequence amount encoding by the programme control. This 64-position rotary drum unit manages printing, control, and document movement functions.

Electric typewriters with OCR keyboards

Olivetti typewriters are available for high or low volume work in any OCR type-font. All have built-in safeguards which automatically prevent misaligned capitals, uneven printing, incorrect spacing, crowded and ghosted characters.



7102ST encoder.



Praxis 48: Olivetti's latest compact electric typewriter.

For full details of Olivetti developments in data processing contact:



olivetti

British Olivetti Limited, 30 Berkeley Square, London W.1. Hyde Park 5011

10. RECOGNITION EQUIPMENT LIMITED

The Electronic Retina Computing Reader*

The Electronic Retina Computing Reader has now been in quantity production for three years and is installed in a variety of industries reading mutilated documents ranging from airline tickets and army papers typed in the field, to authors typed copy.

Thus with well proven design and installation experience the Electronic Retina Computing Reader and associated paper handling equipment affords a unique flexibility for data capture, and sorting.

In brief, the ERCCR offers these features:-

- Non stylised fount capability
- Reads intermixed founts
- Handles intermixed document sizes
- Extreme print quality tolerance
- Acceptance of mutilated documents, staples, tears, dog ears etc.
- Computes whilst it reads
- Fast reading with compatible output
- Transports documents at 200 ins/sec. i.e. maximum reading rate of 2400 characters per second.

The System

The Recognition Unit can be fitted with up to 360 masks to cover a wide range of vocabularies including line printer, typewriter, encoder, OCR 'A' and 'B', adding machine and E13B, all of which can be intermixed on the documents. One or two reading stations can be supplied to permit reading two lines on a document at a single pass.

Each Recognition Unit may have two paper transports associated with it so that the Retina and programme controller could have a Document Carrier and a Page Carrier.

The Document Carrier will handle documents varying in size from $3\frac{1}{4}" \times 3\frac{1}{4}"$ to $8\frac{3}{4}" \times 5"$ intermixed, with paper weights ranging from 45 g.s.m. to card stock also intermixed. Documents are processed at speeds up to 1200 per minute and sorted into pockets supplied in modules of 3, up to a total of 12. Both input and removal of documents may take place whilst the equipment is in operation without affecting the precision feeder or disturbing the neat output stacks.

* Electronic Retina is a trade mark of Recognition Equipment Ltd.

The documents may be carbon backed, dog eared, torn, stapled, crumpled and then flattened or have labels and postage stamps pasted on them.

The Page Carrier will handle pages with irregular and inconsistent line spacing on paper varying in weight between 61 g.s.m. and 120 g.s.m. and in sizes from $4\frac{7}{8}'' \times 3\frac{1}{4}''$ to $14'' \times 14''$. The input hopper and output stacker will both accept a stack of pages 10" thick (approx. 3000 pages of 61 g.s.m. paper). Under programme control the reader can skip horizontal or vertical bands down a page and so selectively read data from a page of information. Also under programme control, characters grouped by class or style can be "turned on" separately or in combinations.

The Electronic Retina Computing Reader, manufactured by Recognition Equipment, incorporates as an invariable element of the system a Scientific Data Systems Incorporated Model 910 computer.

The Programme Controller

This computer provides a flexible capability for 'digesting' data recognised by the Retina. This makes it possible to simultaneously sort, edit, format and record the data in association with peripheral equipment. Typical of the peripheral equipment provided as a part of the Electronic Retina Computing Reader configuration are magnetic tape handlers and a high speed line printer.

The SDS 910 is a medium speed binary computer with a memory cycle time of 8 microseconds. The immediate access magnetic core memory is provided with a basic capacity of 8,192 24-bit words. Each word can contain four characters of 6 bits each. The memory can be increased in modular form up to 16,384 words.

The computer has a library of instructions which, when combined with the available software, provides a comprehensive data processing capability which can be enhanced by the attachment of additional magnetic drum and disc storage units and line printers.

The Electronic Retina

The Electronic Retina Computing Reader system manufactured by Recognition Equipment utilises a unique retinal approach which simulates the human eye. The character recognition portion of the Electronic Retina Computing Reader is known as the 'Retina' because of its similarity to the mechanism and resolution of the human eye. Essentially composed of five hundred and seventy-six cells in an array of twelve cells wide by forty-eight cells long Figure 1, the unit not only senses the character shape but is able to sense the shades of grey associated with a character impression. It is this technique of utilising the 'shades of grey' which is largely responsible for the high performance of the system. The 'shades of grey' ability associated with the Retina, Figure 60 is provided by the fact that an analogue signal is derived from the photo silicon chip behind each cell. This signal provides an electrical signal which is directly proportional to the amount of light falling upon the element. This analogue nature of the signal is retained in the recognition circuitry up to the point of digitally encoding the recognised character. Thus at no time is a false decision made whether or not a point within the character is black or white.

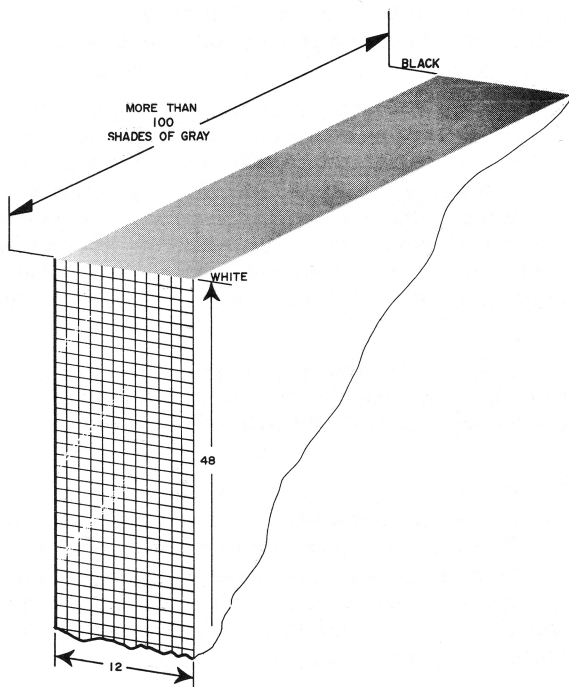


Fig. 60 The Retina

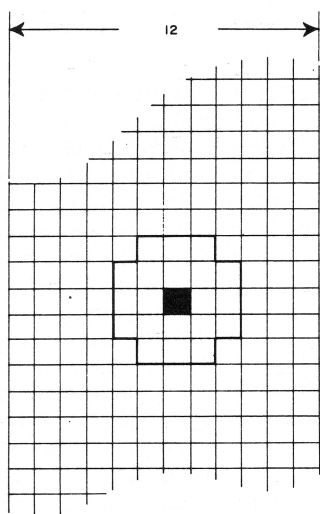


Fig. 61 Amplitude Correlation

The 'shades of grey' ability associated with the Retina provides the capability to 'read' characters of considerable degradation with respect to the background and removes the restriction of the 'zone method'.

The character is also completely contained within the recognition device during the recognition process. This contrasts with a single row of cells or serial scanning systems. This feature permits parallel processing to be carried out simultaneously with respect to the total character shape.

The amplitude correlation feature examines a particular cell in relation to the cells surrounding it according to a mathematical function, Figure 61. If the cell under consideration differs significantly in output from the averaged output of the surrounding cells it is deemed to represent a genuine piece of character information. If it does not differ significantly, it is deemed to be representative of the background and therefore generally does not convey any pertinent information related to the character. If it is significantly different, the cell is either representing a dark portion of the character with respect to the character, or a light portion of the background with respect to the character. In this manner the edges of character shapes can be traced.

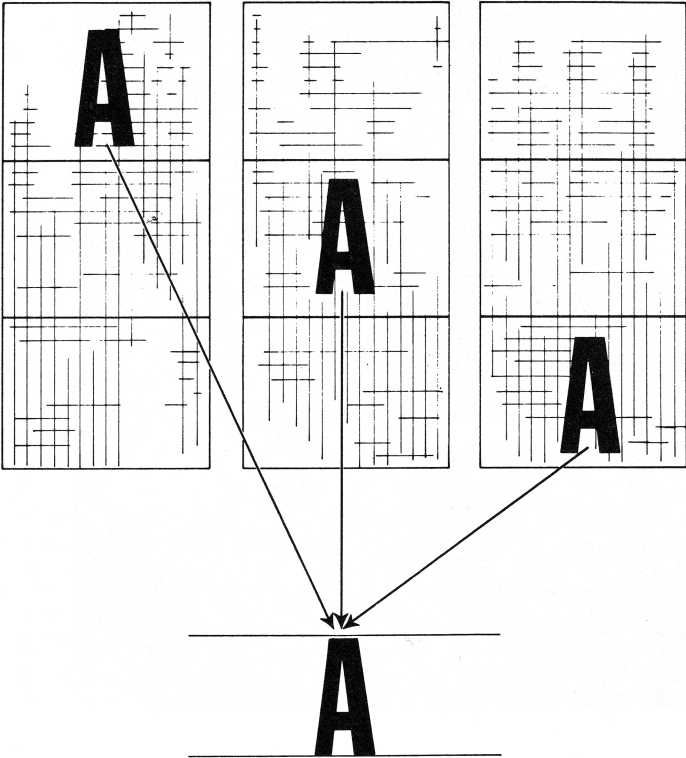


Fig. 62 Misregistration

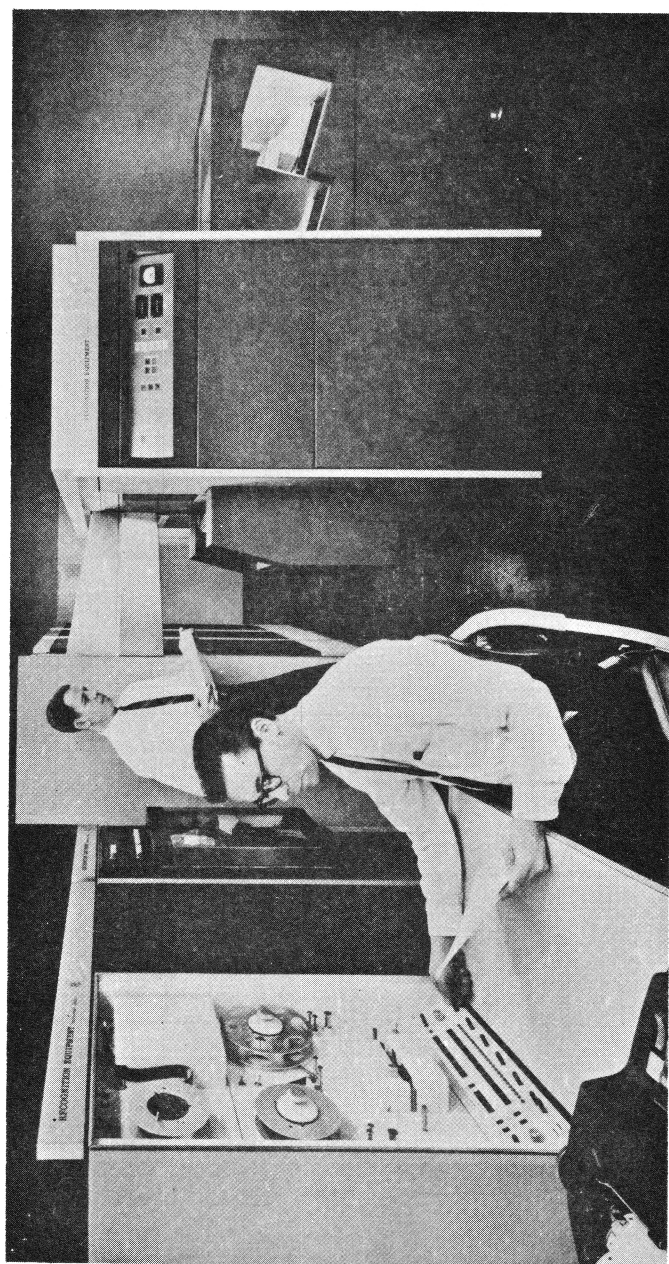


Fig. 63 The Electronic Retina Computing Reader

The recognition process is the key to the reliability of an optical character reader; however, there are certain other vital functions which must be performed in order to ensure that characters can be successfully read. These functions are associated with accommodating the following problems:

1. Line misregistration
2. Character misregistration
3. Line skew
4. Character skew
5. Touching characters

The Retina is arranged to cover an area equivalent to three character heights, Figure 62. In this manner the characters to be read from a line are arranged to fall nominally on the centre portion of the Retina, Figure 62. However, individual lines and individual characters within a line will not invariably appear in this position. They will be either misregistered high or low, Figure 62. The Retina has the ability to examine continuously the position of a character on its face—every fifteen microseconds—in order to determine its exact position. If the examination indicates that a character is high, then the character image is extracted electronically from the Retina in that position and 'normalised' so as to appear at the character masks themselves as if it were in nominal registration. Similarly, if the character is low, the examination determines this fact—virtually instantaneously—and the adjustment made at the point of transfer of information on the Retina to the subsequent analysis and character patterns.

Line skew is accommodated in an exactly similar manner except that skew manifests itself as an iterative misregistration beginning at one end of the line. Line skew is a special case of character misregistration with the difference that the individual characters will be 'twisted' with respect to their nominal shape. This deviation from the nominal is accommodated within the design of the character patterns or masks with which the information imaged on the Retina is correlated.

Touching characters are recognised correctly because the Retina recognition technique intrinsically provides for a character to be recognised whenever it appears in registration with its character mask. This means that no gap is required between characters for timing purposes.

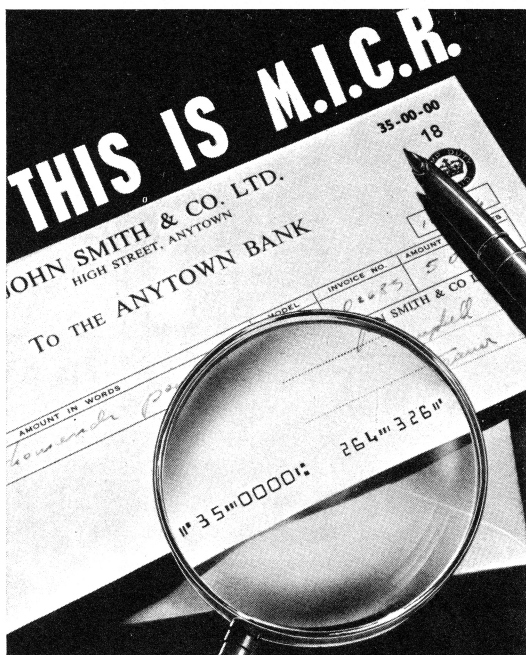
The Electronic Retina Computing Reader cannot be fully described in this handbook but further details can be obtained from:-

Recognition Equipment Limited

Nicholson House

Maidenhead - Berks

Telephone (OMA8)29244, Telex 84466



M.I.C.R. documents printed in E-13B or C.M.C.7 Characters.

O.C.R. cards printed to manufacturers design specification.

Production Control and Stock Order Documents for use on Lector Readers.

Payroll transaction documents in flat sheet or continuous form for use on O.C.R. equipment.

A specialist service based on experience.

WATERSTONS

GEORGE WATERSTON & SONS LTD.,

P.O. Box 72, Warriston Works, Edinburgh 7.

5 Morocco Street, London, S.E.1.

Waverley 4471
Hop 7387 & Hop 0064

9374

SECTION 7

Cost Comparisons

COST COMPARISONS

Data Handling procedures include:

- (a) generation of the source document
- (b) transcription of the information if this cannot be read easily by keyboard operators because of poor layout or quality of writing
- (c) operation of key sets or automatic devices such as mark reader/card punches
- (d) reading into the data processing machine (computer)

Other chapters have stated that the aim in data handling is to eliminate stages (b) and (c). Eventually, perhaps stage (a) might also be eliminated, by the development of detailed information models in a total system, constantly updated by variations, which might be keyed directly in to the computer.

This chapter does not consider the development of data capture by means of key sets attached through multiplexors to a central processor and main file storage. This is a technique of which there is little practice at the present moment, which requires multi-programming facilities and special verification techniques with possibly video displays or audio responses.

This section is concerned with a comparison of the various alternatives insofar as this is possible.

The standard method of data preparation against which comparisons can be made is:-

Keypunching and verifying cards—about which much is established and on which various factors can be applied by individual users to allow for special location or work difficulties or advantages.

Alternative methods include authentic character recognition techniques on:-

Turnaround documents including documents with preprinted code lines.

Mark scanning as distinct from mark sensing.

By-product from Keyboard operation e.g., tally rolls.

Typing and character reading.

Turnaround documents include counterfoils from, for example, computer printed returnable insurance renewal papers, rate demands, public utility bills and returnable hire purchase instalment advices. These eliminate (b) and (c) and undoubtedly offer the greatest economics in these situations. Nevertheless applications are subject to limitation. They can operate only where the data is known in advance and can be pre-recorded or where machine-readable print is employed on preparation of the source document. The imprinting of bank, branch, customer and cheque serial numbers on cheques is also in this category, but the variable amount requires the equivalent of key punching. Pre-

printed documents and punched cards are forms of turnaround documents to which a limited amount of modification can be introduced by marking in predetermined spaces to eliminate (b) and (c).

Mark sensing simply substitutes a machine operation for the human operation of key sets and thereby eliminates (b) only.

In all cases, however, there are large costs involved in setting up files of data and in keeping them up to date.

In an operation of a computing system, the cost of preparing data is often a higher proportion of the total than might be first imagined. More work of every kind will be applied to computers when cheap, automatic reading of documents at high speeds becomes possible. The transfer of data on to cards and tapes and then checking for accuracy is slow and costly in time and materials. The data transferred cannot be more accurate than the original and errors may slip through. Relatively cheap automatic reading machines are now available for business purposes: conceivably they can be used for scientific work where much repetitive work is recorded and computed.

COST COMPARISONS

It is not possible in a Handbook of this size to cover in any detail all the infinite variety of factors that affect the cost, for these may vary so much between different systems that to give any example tends to become a purely theoretical exercise. Furthermore, a system that shows a saving for one user may not show a saving for another of the same industry, because of differing factors. Particularly is this so for turnaround systems which can be so highly rewarding.

Two cost comparison have been made.

The first shows as a rough guide the point at which it is profitable to use a reader instead of keypunch operators in a complete turnaround system.

The second is the alternative system of converting raw data by typing and character reading instead of keypunching and verifying cards or paper tape. Typing and reading is being practised widely in the U.S. and it is on that evidence that the costs have been estimated. There is NO information from any BRITISH SOURCE to CONFIRM the findings.

What is in fact considered is the conversion of mainly alphabetical data as for names and addresses, where the practice is to type once only and to sight check the hard copy before automatic reading.

For numeric data it is not quite the same approach. It is possible however, to achieve a high degree of accuracy by typing twice and reading both copies into the computer (using character recognition techniques) and writing the data to two magnetic tapes, which are then computer-compared to 'verify' the data. Even this lengthy procedure might not seem to make an unfavourable comparison with keypunching but its effective use cannot be realised unless very large volumes of data are to be converted.

In each case comparisons of costs are taken from the point of receiving raw data and converting it through a keyboard to writing it to magnetic tape.

KEYPUNCHING CARDS OR PAPER TAPE

In assessing the costs of keypunching the following assumptions have been made—

1. That the punch rate depends upon a number of different factors, amongst which are: -
 - (a) the layout of media and the readability of the printing and/or writing.
 - (b) the mixture of data, numerical and/or alphabetical.
 - (c) the variety of media
 - (d) the size of batches of work
 - (e) the number of decisions to be made by the operator.
2. That the key depression rate averages out at 8, 000 per hour using a full alphanumerical typewriter-style keyboard. Verifying rates are normally 10 per cent higher than punching which has been taken into consideration in setting the average.

An average of 7, 000 key depressions per hour for cards might be taken by many people but here the comparison anticipates alternative methods to cost less and a better keypunching performance makes the comparison more conservative.
3. That within all cards there is an average of 50 columns punched and that all are verified, with no gang punching. Effective column rate is therefore 3, 500 per hour.
4. That 5% has been added to the cost for a correcting procedure.
5. That the cost of writing correct cards away to magnetic tape by computer is based on—
 - (a) Card reading speed of 500 per minute.
 - (b) Transfer rate of 30, 000 characters per second.
 - (c) Computer cost of £30 per hour.

Rates will vary considerably in different parts of the country. Therefore a proforma of headings is given to enable the reader to calculate his own costs (Appendix 1). The following schedules are based on London prices during the first part of 1966.

OPERATOR/MACHINE COST PER WEEK

	£	
Staff Wages	14. 2	
Tax	0. 6	
Training	0. 2	
Machine	3. 0	
Office Space	1. 5	(45 sq. ft. including service)
Supervision	1. 6	(1 Supervisor to each 12 operators)
Cards	1. 0	(1750 cards)
	22. 1	
Overheads at 40 per cent of wages	5. 7	
	£27. 8	

Much the biggest variant in costs per 1000 cards is the effect of maintaining sufficient staff and equipment to deal with peaks of work which for example may occur during any day or during any week or at the end of each month. The following shows the difference in cost according to loadings in hours per week.

Hours per week loaded	20	25	30
No. of Cards per operator	1600	2000	2400
Cost per 1000 cards	£17. 7. 6	£13. 18. 0	£11. 11. 6
Add allowance for correction	17. 6	14. 11	11. 7
Cost of computers reading 1000 cards and writing mag } tape.	£ 1. 0. 0	£ 1. 0. 0	£ 1. 0. 0
Total to mag. tape per 1000 cards punched and verified }	£19. 5. 0	£15. 12. 11	£13. 3. 1

NOTE: No figures are given in excess of 30 hours a week since this is the generally accepted effective working hours obtainable.

OPTICAL READING AND WRITING DATA TO MAGNETIC TAPE

In assessing these costs the following have been assumed:

1. That the Character Reader is not on-line to the computer but that it has an output of $\frac{1}{2}$ inch magnetic tape.
2. That the Character Reading speed has been assessed at 150 characters a second to allow for all contingencies although theoretical reading speeds are much higher. This gives the equivalent of reading 10, 800 records per hour, each record consisting of 50 characters.

<u>Reader Costs per week</u>	£
Staff Wages etc.	45
Rental	350
Office space	9
Power and ancillary expenses	10
	414
Overheads at 40% of wages	18
	£432

NOTES: The rental of the reader is based on quotations obtained.

TURNAROUND SYSTEM

It is therefore possible in a complete turnaround system to make a very broad assessment of the point at which it is economic to use a character reading machine rather than to employ operators on key punches.

The comparison is concerned with the system at the point of re-entry of the data, as for example in the case of premium renewal notices and hire purchase payment notifications.

By dividing the cost of the key punch operator into the cost of the reader the break-even point is to the advantage of character reading when the load exceeds the capacity of 15 key punch operators.

TYPING AND OPTICAL READING

Consideration is given to the conversion of mainly alphabetical information as required in building up a file of names and addresses. Assumptions are made as follows:-

1. That the key depression rate is influenced by the same factors which are listed as affecting keypunching except that names and addresses would often be taken from good copy.
2. That the data will be typed only once, but that each document will be sight checked. Checkers are used in the proportion of 1 checker to 3 typists.
3. That the rate of typing is taken as 40 words a minute (normally taken as 50) with a word length of 5 characters. Normal electric typewriters are used except that the keyboards may incorporate a second bank of numeric keys in the same order as the number keys on a keypunch, with extra keys for special symbols.
4. That the equivalent key depression rate is 12, 000 per hour.
5. That for comparison purposes the cards equivalent is assessed at 50 columns. Names and addresses can be assumed to total 100 characters or two cards equivalent.
6. That an allowance of £2 per 1000 cards equivalent is made for subsequent corrections.
7. That reading is 10, 800 records each of 50 characters per hour which in a week of 40 hours is 432, 000 or the output of 60 typists working to a capacity of 30 hours each.

<u>Typist and checking costs</u> per typist week	£
Staff typists' wages etc.	13.0
checkers' wages etc.	4.0 (1-3 typists)
tax	1.0
training	0.3
Machine	1.4
Office space	1.7
Supervision	1.0
Paper	0.5
	£22.9
Overheads at 40% of wages	6.8
	£29.7

Reader costs are as previously stated—£432 per week

Hours per week loaded	20	25	30
No. of cards equivalent typed per typist week	4800	6000	7200
Cost of typing and checking 1000 cards equivalent	£ 6. 3. 9	£ 4. 19. 0	£4. 2. 6
Cost of reading per 1000 card equivalents			
At 20 typist level	£ 4. 10. 0	£ 3. 12. 0	£3. 0. 0
At 40 typist level	£ 2. 5. 0	£ 1. 16. 0	£1. 10. 0
At 60 typist level	£ 1. 10. 0	£ 1. 4. 0	£1. 0. 0
Total cost of typing checking, reading per 1000 card equivalent and writing data to mag. tape			
At 20 typist level	£10. 13. 9	£ 8. 11. 0	£7. 2. 6
At 40 typist level	£ 8. 8. 9	£ 6. 15. 0	£5. 12. 6
At 60 typist level	£ 7. 13. 9	£ 6. 3. 0	£5. 2. 6
Plus the cost of correction	£ 2. 0. 0	£ 2. 0. 0	£2. 0. 0
At 20 typist level	£12. 13. 9	£10. 11. 0	£9. 2. 6
At 40 typist level	£10. 8. 9	£ 8. 15. 0	£7. 12. 6
At 60 typist level	£ 9. 13. 9	£ 8. 3. 0	£7. 2. 6

Speeds of typing will vary very much for the reasons previously mentioned. To these can be added incentives, the effectiveness of supervision, morale, the state of the labour market and working conditions. A significant advantage is that the operator can check a 'hard' copy and correct by blocking out and continuing typing.

COMPARISON OF COSTS OF KEYPUNCHING CARDS AND TYPING AND OPTICAL READING

A comparison of the above with the cost of punching and verifying 1000 cards and writing to magnetic tape previously shown—

Hours per week operators loaded	20	25	30
	£19. 5. 0	£15. 12. 11	£13. 3. 1

indicates the potential of typing and reading to reduce considerably the costs of data conversion, particularly for large volumes of mainly alphabetical content.

In the example given above the extremes in cost savings have been taken and it should be borne in mind that this can only be a guide to the savings that can be expected.

Few comments have been made regarding the relative costs of systems based on the reading of tally rolls, or the use of mark scanning (magnetic or optical), or the use of a by-product, such as papertape, from another process.

Because of the difficulty in covering all the possible factors that may affect the cost only a relatively few have been considered. Therefore a list of the main factors that may be expected to vary as a result of the adoption of a new system of data transcription is given at the end of this chapter, together with a more detailed list of questions that may need to be answered in working out a cost comparison.

APPENDIX 1

COST HEADINGS

PRIME COSTS

OFFICE SPACE

RENT

RATES

HEATING LIGHT AND POWER

CLEANING

MACHINES

RENT

DEPRECIATION

MAINTENANCE

STAFF

WAGES

NATIONAL INSURANCE AND PENSIONS

SELECTIVE EMPLOYMENT TAX

HOLIDAY CREDIT

TRAINING

AMENITIES

CARDS OR STATIONERY

SUPERVISION

ON COST

MANAGEMENT

OTHER COSTS

APPENDIX 2

THE SYSTEM

Will the changeover involve extra work for some classes of employee?

If so to what effect and at what cost?

Does the cost of preparing the original media have to be considered?

Is it a by-product such as a tally roll or punched paper tape?

Does it take the place of a previous form of media?

Is the original media used as input to the machine?

What is the cost of transporting media to machine?

Is transcription necessary?

If so, is it centralised or decentralised?

What is the cost of rectifying errors?

Is any preparatory or additional work required?

Can all exceptions be handled?

If so, at what cost and loss of efficiency?

Is it necessary to maintain a dual system during installation?

What would be the cost?

Is there extra work during changeover?

Are there any additional checks and other supervisory work required?

Are there any forms, equipment, or supplies made obsolete by change?

If so what is the loss?

What is the cost of redesigning stationery?

Are there special papers required?

Will there be 'Turnaround' documents?

If so what is the cost of original preparation?

Is there to be verification of documents to be read by the machine?

Are there any extra handling costs?

Will format and presentation of information to customer and/or internally be improved or deteriorate?

What are the indirect advantages to other departments, for example to salesmen in this field, or to stocktakers?

What value can be placed on these?

OTHER OFFICE STAFF & EQUIPMENT (excluding Character Reader)

Will the administrative system change?

If so, what is the cost?

Is any extra furniture required?
Can existing furniture be used?
If not, what is the cost of replacement?
Is there any alteration in office stationery or supplies?
Is there any difficulty in obtaining supplies?
Will the number of secretarial and administrative staff be altered?
If there is an increase what will be the cost?
If taken as a reduction can the surplus be employed on other duties?

BUILDINGS

Will more, or less, space be required?
If so how much? — and at what cost?
Will heating, lighting, power or water charges be increased or reduced?
Is the building suitable for the proposed system?
Are any building alterations required?
If so at what cost?
Is any or extra air-conditioning required for the reader?
Has sufficient time been allowed for the proposed alterations?
If not, what will be the extra cost of completing on time?
What is the cost of over-running deadline?
If another building is required what will be the cost?

MACHINERY

Should the character reader be purchased or rented?
What are the differences in cost between existing machinery and the proposed machinery?
If renting, are maintenance and repair costs included?
If buying, is loss of interest on the capital included?
Is a maintenance contract required?
If so, are repair costs included?
Are there Income Tax allowances on the capital cost of the machinery purchased?
Are the rates of depreciation higher or lower?
Is more or less floor space required?
Are there any special environmental conditions?

What will be the extra cost of special stationery, ribbons, ink etc for the proposed machine ?

Will operators be required to be of higher grade?

Will there be a loss on the machinery replaced?

Are there to be changes in ancillary equipment?

PERSONNEL

Are there any changes in size of staff?

Is a higher or lower calibre of staff required?

Are any changes necessary in the type and period of training?

What is the cost of re-training existing staff?

Are there any shortages in supply of suitable staff?

What will be the effect of the change on morale?

Will there be any changes in terms of employment and/or working conditions? If so what is the cost?

Section 8

Systems Techniques

Some Practical Applications

SYSTEMS TECHNIQUES. SOME PRACTICAL APPLICATIONS

This Section describes some examples of systems actually in use, and the list is drawn largely from the United Kingdom. Two systems show the application of the E13B and CMC7 MICR type fonts respectively; the remainder exhibit examples of Mark Reading and OCR techniques.

The complete range of applications described covers a variety of data media, ranging from small discrete documents through large page documents to continuous-strip tally rolls. Where data is input to a Computer, there is, in general, no need to sort the media on which the data is presented but, in banking and giro etc. operations, sorting is usually essential; this fact becomes apparent among the example systems involving small documents. A major distinction can also be made among the OCR and Mark Reading examples, on the basis of whether the system is 'turn-around' or 'open-ended' i.e. whether the data to be read is printed by the computer with which the reader is associated or derived from sources in no way connected with that computer.

It is essential to realise that, in a system providing data collection by character recognition means, a proportion of the data will be illegible to the reading device. Readers are designed to ensure virtually negligible error rates but a proportion of rejected data must always be legislated for. In MICR, where printing and encoding standards are accepted as having to be high, a document reject rate of as little as one per cent is not unusual for the small documents in question. In OCR, where poorer print standards have to be tolerated the reject figure may be substantially worse; and this is likely to be the more so in the 'open-ended' class of application.

In some practical cases, it has been found viable to re-read the rejected data, with a consequent reduction in the reject residue, but many systems practice immediate transcription of the unreadable material after a single reading operation. In document sorting circumstances there may be rejects after each pass, and the problem of handling these items at different stages of processing is very often minimised by reading in the first pass the complete data item on which subsequent sort passes are to be based; this technique is readily made possible by connecting the sorter/reader on-line to a computer.

Apart from the obvious one of punching a card or paper tape record in respect of rejected data, various methods of transcription are adopted for making use of the character reader itself. The original document can be resubmitted to the reader after erasure and re-encoding of the data in the same position, re-encoding in a new position or re-encoding on an attached correction slip; other alternatives are the placing of the document in an encoded envelope or the encoding of a substitute document. Some of these methods are as much applicable to the correction of encoding errors as to the transcription of unreadable data.

In general, the whole operation of handling reject data becomes more acute as the amount of data per document increases, since the segregation of documents carrying unreadable characters then becomes less

and less valid as the criterion for data transcription. For this reason, in page and continuous-strip tally readers, means are very often provided for the marking of unreadable items for subsequent inspection, or even for their immediate display to an operator for human interpretation and keyed entry.

WESTMINSTER BANK LIMITED

Early History

The first fully operational use of electronic data processing techniques commenced within the Westminster Bank in the second half of 1962. It was strongly believed, during the prior planning period, that the really big benefit of any EDP system within a Bank must include automatic magnetic ink character recognition cheque reading and sorting techniques as an integrated part of that system. So, for that very reason, the first IBM computer to go into operational use was delivered with a sorter/reader designed to interpret the already well proved and established E13B magnetic type-font. Since that time there has never been any cause to regret the choice of MICR within a cheque handling process and at the time of writing the number of cheques daily handled peaks to over 600, 000 and is weekly increasing. With such background it is a pleasurable task to offer this contribution of overall experience with a MICR application, from a large user's viewpoint. Fig. 63 shows a typical cheque sorting configuration.

Necessary Requirements

A few comments regarding the preliminary problems and early researches may be considered appropriate, as when designing a system employing magnetic character reading, there are several essentials to be considered which may be briefly summarised in the following paragraphs.

- (a) The document will be accelerated from rest very rapidly and its physical composition should be sufficiently strong to accept this acceleration without buckling or tearing in the transit mechanism of the sorter.
- (b) The magnetic characters must be distinctly recognisable within the logic of the sorting mechanism. They must be reasonably easy to print on the document and a high degree of quality control is most desirable during the printing operation.
- (c) The document handling equipment must be able to accept some latitude from the standard specification of the position of the code line, and it should be able to over-ride some very small imperfections in the printed character.

The Bank of America, with the co-operation of the Stanford Research Institute of California, had already spent many hundreds of thousands of dollars developing a functional type-font, resulting in a character specification known as E13B and, with the help of the valuable experience already gained by our American Banking colleagues, and the benefit of considerable experiment and research carried out under the direction of the Electronics Sub-Committee of London Clearing Bankers, by 1962 specifications were sufficiently determined to satisfy all the conditions required within paragraphs (a) and (b) and proven equipment was available to deal with (c). Being then in a position to accept

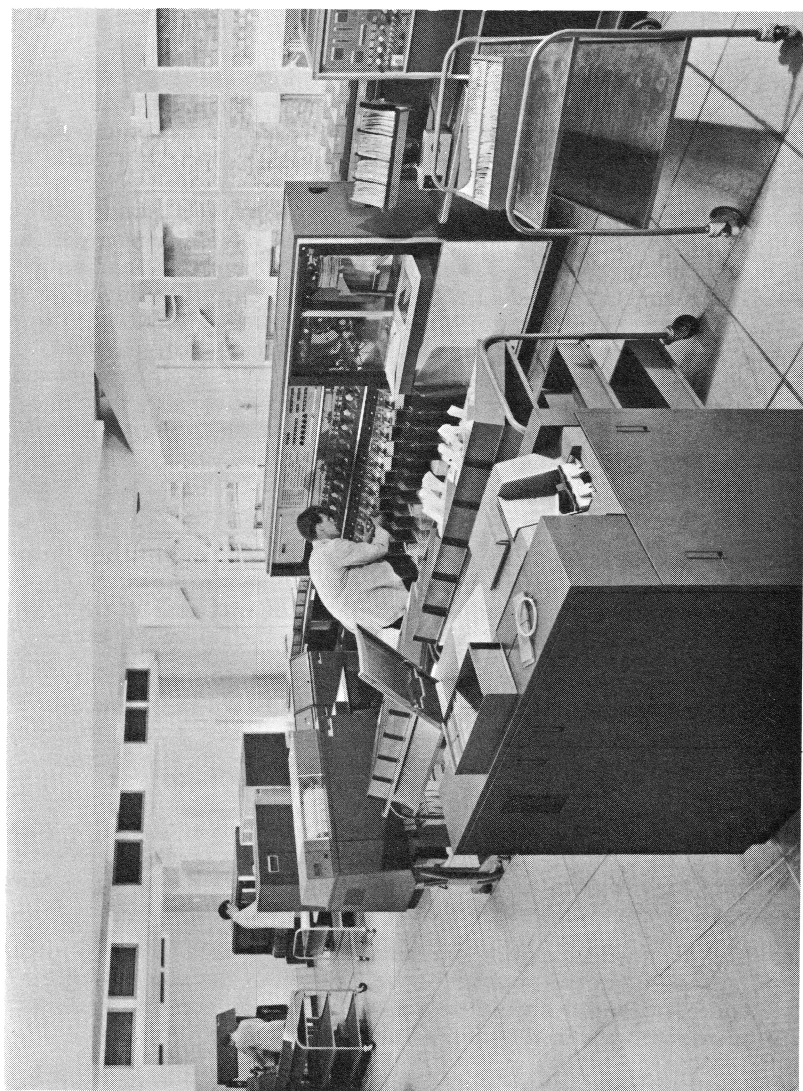


Fig. 63

fully the advantages of direct magnetic character reading, the Bank was able to go 'live' during the second half of 1962 with machinery already proved in U.S.A. and available for early delivery in the U.K.

The detailed specification of the E13B type-font is fully set out in the handbook published by the Committee of London Clearing Banks entitled 'Requirements for Automatic Cheque Processing', and the physical limits in this specification regarding the printing and location of the code line along the lower edge of the cheque, are set to a very fine tolerance. Although initially the printers found it difficult to hold these limits they quickly overcame production problems, and the judgment of that Committee and the implementation of its requirements by the printers undoubtedly contributed greatly to the ultimate success of the application.

A formidable task was presented to the Bank's Stationery Office, and to the cheque printers, in order to put into circulation sufficient magnetically encoded cheques on the new stiffer cheque paper (specification CBS) to make automatic document handling a practical proposition, and not just an interesting experiment.

MICR Fully Operational in U.K.

Early 1963 saw the Bank handling a large proportion of its City business, comprising the large Lothbury, Lombard Street, and Threadneedle Street Offices, employing full MICR techniques. The involvement in this 'deep end' approach was not without its problems and called for strong nerves, but the solution of the various difficulties gave early and invaluable experience and permitted the Bank to go ahead with confidence to further its MICR programme.

The In-clearing Completely Automated

So that the whole of the IN-Clearing may be sorted automatically, the cheques of all Branches, whether serviced by the computer or not, are now pre-printed in magnetic characters with the nationally agreed sorting code. Additionally, the cheques of those Branches already on the computer have the serial numbers and customers' account numbers also pre-printed.

To complete the process the cheques still have to be encoded with the amount of the entry before being acceptable to the automated system and this work is done as a by-product of a normal listing operation with specially developed 'encoding machines'.

The full application of MICR techniques in the Banking context did not spring up overnight but from small beginnings the whole system very quickly grew. Now, in the year 1966, the whole of the In-Clearing is automatically sorted, listed, and in the case of those Branches on the computer, directly applied to the appropriate accounts.

Fig. 64 shows a specimen cheque carrying a complete line of encoded information.

Experience in Operation

It was soon found that the sorter/readers had considerable tolerance regarding the condition of the document and location of the code line. It was originally feared that cheques which had been folded in pocket

0-1

16th July 1966 60-00-01T

WESTMINSTER BANK LIMITED

ANYTOWN BRANCH
12 HIGH STREET, ANYTOWN, BERKS.

Pay to the order of

One hundred and three pounds **£103-18-5**

Eighteen shillings and five pence

JOHN ENGLAND

John England

WESTMINSTER BANK LIMITED

W 22790018

Customer's Account Number

Amount

E 13 B

Code Line

Cheque Serial Number

Bank & Branch

Sorting Code

Customer's Account Number

Amount

Fig. 64

books or creased in handling, might jam up the sorter. Thanks partly to the excellence of the new paper specification this was not so, although the occasional mechanical stoppage does still occur. This condition is, however, very easily cleared and this facility must be inbuilt in any good sorter. The presence of steel staples remaining in the cheques during the first pass of the sorter presented some considerable embarrassment and a certain amount of damage to equipment was experienced. If the staple is located anywhere within the code line it is certain to corrupt the entry and the item will be passed into the reject pocket, although there is no possibility of a wrong amount entering the system. To deal with this hazard a small team of girls was forthwith established to remove the staples prior to processing; the necessary device for this operation was found at one of the City stationers costing about 2/6d. This non-productive team has now been disbanded due to vigorous customer education and strict instructions to Branches to examine their Out-Clearing to ensure that all staples have been removed.

At this point it is appropriate to emphasise that no matter how sophisticated may be the magnetic reading circuitry and logic, and no matter how ingenious may be the sorting mechanism, the user is courting almost certain failure if he has not paid very close attention to the physical characteristics of the paper and the document to be so processed. Nothing is more dis-spiriting than a snarled-up sorter and those potential users, who are even now investigating the handling of tissue-like forms are strongly recommended to re-consider their problem.

Within the MICR application, 'piggybacks' were in the early days a source of annoyance. These, as implied, consist of one document going through the sorting gate on top of the other. The magnetic characters of one are correctly read, but the obscured document is with certainty directed into the wrong pocket and is not read. Static electricity or creases causing adhesion may create this trouble but with the more advanced equipment now available this occurrence is almost entirely eliminated.

A small proportion of cheques may pass into the reject pocket due to faulty printing, smudges, or voids occurring within the E13B code line. In such cases it is quite impossible to hold an inquest some days later as the offending item must be cleared, with the others, on the same day (there can be no 'tomorrow' in a Banking clearing operation). A Cross-field MICREX electronic comparator has made possible an on-the-spot examination of rejected items and with experience it is possible to track down very quickly the faulty character. This is recorded in a 'Black Ledger' detailing the cause of the reject and the printers name. This process has been equally welcomed by the printers who are able to back up their own very careful quality control with reliable user experience.

The Processing Routine

As the main purport of this article is to discuss the overall function of the MICR application within the Bank, it is not required to explain in detail the actual sort pattern of the cheque handling process which, with its inbuilt accountancy tests and precautions, would justify a lengthy paper. The cheque reading complex does, however, consist of five IBM

type 1419 sorter/readers, each linked to a 1401 control computer commanding the sort routine. Each of the sorting units is reading and simultaneously sorting magnetically encoded cheques of mixed sizes at approximately 1, 100 items per minute on each pass, and it is estimated that the proportion of cheques so processed represents about 60% of the entire debit throughput.

In the case of those many Branches already serviced by the computer the cheques drawn thereon will carry the whole case history of the transaction in magnetic characters (viz: serial number, bank and branch sorting code, account number and amount). This information, captured during the sorting run, is simultaneously placed on magnetic tape for application to the appropriate account in a separate computing operation.

Computer Branch cheques are also fine-sorted down to an alpha-numeric sequence in order that, when they arrive at the Branches on which they are drawn, they may be more easily scrutinised for authenticity of signature and technical correctness, a final and important stage which can only be done by human examination. Should any cheque not comply with these requirements a simple procedure deletes the transaction from the computer file. Thus, human judgment may always over-ride an electronic process.

Operator error can occur when the cheques are being magnetically encoded for amount, and such errors are deleted by a magnetic pencil which, when passed through the incorrect marking, completely voids the reading facility in the amount area of the code line. Adjustment of such errors is effected by substitute documents and reconciliation adjustments.

It has been found difficult to determine the percentage error rate directly attributable to faulty reading of the magnetic character, as the reading logic of the sorter/reader will not inject suspect information into the system. If in any doubt the document is passed into the reject pocket. This would include documents which had received rough treatment prior to their arrival into the system and where the magnetic code had been corrupted by heavy folding or tears. The figure due to these causes is about 1.6% of throughput and is improving.

The Credit Clearing

Most of the remarks in this paper have been directed to the DEBIT-Clearing. Magnetically encoded credit slips of standard specification and size are already included at the rear of the cheque book but automatic handling of mixed credit items by MICR offers problems of a different nature to that of the cheque clearing operation, as at present the Banks have no control of the size and shape and paper specification of credit documents which may arrive from many sources.

The essence of rapid and expeditious handling in a sorting operation is standardisation of the document, physically and in respect of the magnetic coding. For these reasons alone the CREDIT-Clearing is not yet automated, and certainly not for any purely technical reason. This position is, however, under close review and it should not be long before a large amount of the physical processing of the credit side of the Banks is automatically applied with identical MICR techniques. In this application the wide possibilities should also be considered for handling the large volume of credit entries arising from transactions within the con-

sumer industries (e.g. electricity, gas, water). Here the Banks will be able to hand over, daily or periodically, magnetic tapes representing payments or dues, read directly off the credit document for injection into the concern's own computer.

2. BANKGIRO, SWEDEN

Introduction

All the commercial banks in Sweden participate in the Bankgiro operation, two of which also represent the savings banks and the agricultural credit banks. Altogether 3, 800 bank offices, spread throughout the whole country, are linked to the central system.

Three types of transactions can be made: transfers from a cheque account maintaining a giro number to a beneficiary holding a giro number, payments in cash from an account maintaining a giro number to a beneficiary not having a giro number, and deposits into an account with a giro number by someone who does not have a giro number.

The remitter delivers the transaction orders to his bank office. The bank withdraws the total amount from his account, returns to the remitter a stamped copy of each payment order and sends the originals to the Giro Centre (Bankgirocentralen).

The Giro Centre—which is situated in Stockholm—has as its primary task to receive giro orders from the bank offices, to sort the orders, to check and microfilm them, to advise the beneficiaries and their bank offices and to take care of the clearing between the banks in respect of giro material. The Centre is responsible, furthermore, for the bank's joint information activity in respect of giro operations, designing and printing of forms, catalogues, brochures, work instructions, etc. On the other hand, the Centre does not deal with the bookkeeping of the remitters' or beneficiaries' bank accounts. All the 'money' which flows into the Centre in the morning therefore leaves in the afternoon, i.e. all transactions are concluded during the day they arrive at the Centre.

The System

Transfer and deposit documents, when received in the Giro centre, are encoded with the amount and the giro number of the beneficiary in the CMC7 MICR font, and are then read into ICT type 1500 processors via 8500 sorter/readers connected on-line. All the MICR information is thereby transferred from the orders to magnetic tapes and, at the same time, sorting of the orders is commenced in accordance with one of the six digits in the giro number. The magnetic tapes are sorted to numerical giro number order and are checked subsequently with a master tape containing instructions for closed accounts, etc. After the initial input pass, the orders are sorted off-line by the remaining five digits in the giro number. Advantage is taken in the on and off-line passes, of facilities for early out-sorting of high volume accounts. Further, the tapes derived from the several input streams are merged together on a single tape, and the latter is then run with a name/address file tape, whereby credit advices to the beneficiaries are printed out. One final check is made visually that the beneficiary's name on the payment order tallies with the name on the credit advice to the holder of the credit giro number. Thereafter, the credit advices and the associated payment orders are microfilmed and placed in sealed window envelopes for mailing. The material is ready for posting as a rule before 6. 30 p.m.

After the credit advices have been printed, the consolidated input tape is rearranged according to bank offices, so that a credit voucher for each bank office can be printed out. The credit voucher is sent to the bank office of the beneficiary for bookkeeping or—if the bookkeeping is done centrally in the bank—for their information. If the bank's book-keeping in respect of cheque accounts is effected centrally, a magnetic tape is handed over to the bank containing the data which is required for the entries to be posted to the beneficiaries' accounts.

As regards payments to beneficiaries who do not maintain giro numbers, the Centre's work consists mainly of providing the orders with the amount written by a cheque protecting machine and the period of validity as well as, for the remitting bank's account, making a record of the orders issued and cashed. This is done by means of punched paper tapes.

Configuration of Processing Equipment

The layout of the installation is shown in Fig. 65. Four ICT 8500 sorter/readers are connectable to one of two ICT 1500 processors through a switching unit. The system is based on serial magnetic tape processing, for which two six-deck 30 Kc/s sub-systems are provided for processor connection via a second switching unit. A variety of output tape formats is provided for compatibility with member bank systems.

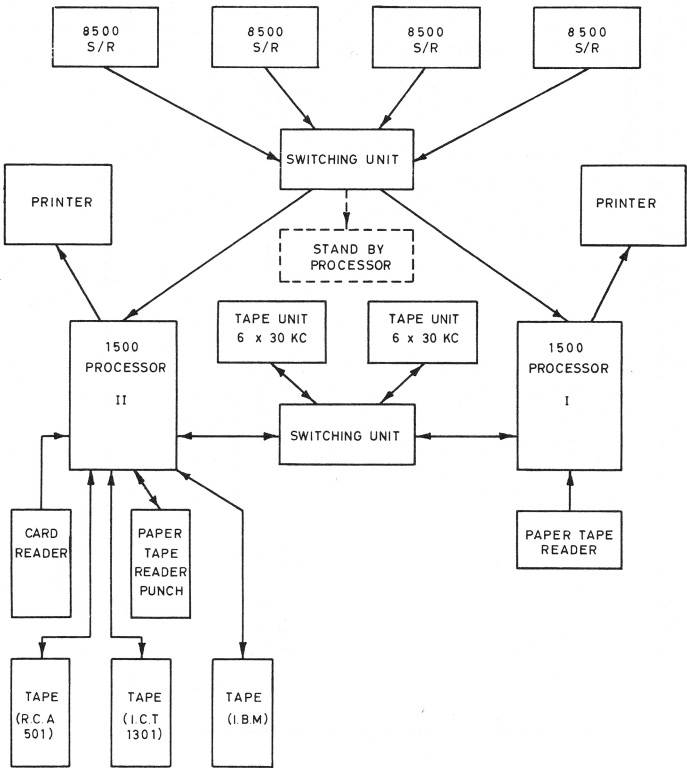


Fig. 65

BANKGIRO GIRERINGS/UTBETALNINGS-AVI		Nr 29094	
Faktura 965435 Kr 4.550:-- Från bankgiro nr 99-1234 Betalningsavvärdare AB Norhammar Box 87 NORHAMMAR		93-7586 AB Järn & Stål Kvitens vid kontantutbetalning PROVBLANKETT	
År bankgironummer ifyllt här nedan, göllgöres beloppet mot bankgironumret svarande bankräkning. År det ej ifyllt, kan beloppet, certifierat här ovan, lyftas av mottagaren i bank. Till bankgiro nr Betalningsmottagare (vid kontantutbetalning även adress)		Avvärdarbankens stämpel MELLANSVENSKA BANKEN 14 DEC. 1965 4199 1	
BG 10 1964.6. D		937586 10000455000 Giro No. Amount	

Fig. 66

Document Aspects

During a normal day, about 50,000 transfers/deposits will be processed and, at the beginning of each month, the number of documents handled approximately doubles. The system started in November 1959, and the number of transactions has been increasing steadily since that date.

The documents concerned are largely of the following size, 210×99 mm, or in some cases $210 \text{ mm} \times 4 \text{ inches}$; and the system never processes documents in mixed size batches. At the machine reading stage, the paper quality of the documents is in general very good.

More than a year before installing the ICT 8500 sorter/readers, all new giro forms were printed on the agreed paper for automatic MICR processing, so that there were no problems at the change over to an automated system.

The location of the CMC7 code-line is seen from the specimen document in Fig. 66. The information consists of a credit giro number, an amount and a control code. The reason why the code-line is so far from the leading edge is to allow space for the bank stamp which is to be applied on the lower right corner of the document. The symbol sequence follows the recommendations agreed by the Swedish Banks Association. There is no preprinted MICR-data on the documents.

To encode the MICR characters on the documents about 50 IBM 1206 inscribers with Addo adding machines are used. Since the giro number does not include any check digit, a very strict control of the inscribed data has to be done. Summated amounts are agreed against batch totals.

Fig. 67 shows a sample computer-printed credit advice.

Operational Experience

The system has been working since June 1965, and in full operation since November 1965. The documents are generally run in batches of about 10,000 at a time. The nominal rate of 1,200 documents a minute cannot quite be achieved, because operation has to be stopped at certain intervals for printing out control amounts, etc. The practical rate amounts to 1,000 documents a minute.

At the start, before the encoders and the sorter/readers were tuned in with one another, the reject percentage was relatively high but has now decreased to lie steadily between 0.5-1%. No further attempt is made to rerun these rejects and they are treated manually.

BANKGIRO		UPPGIFT OM INSÄTTNING	
BANKGIROCENTRALEN Belopp 725,43 1060,50 4550,00 10860,85 12900,00 17465,00 * S: A KRONÖR 47561,78		BANKGIROCENTRALEN Bankgiro nr Bank Bankräkning nr Antal avier 93-7586 NSB 9999 9876543 6 Betalningsmottagare	
AB JÄRN & STÅL BOX 123		VÄSTERSTAD 67	
Till räkningen kan även följande betalningar ställas till		STÅLBOLAGET	

DEC 1 5 1965

BG 40 (7) apr 1965

Fig. 67

3. WEST MIDLANDS GAS BOARD

A data capture system, based upon computer-produced bar-markings on customer consumption and hire-purchase accounts, has been in use for two and a half years. Four English Electric Leo Marconi 'Lector' document readers now handle each day between 25, 000 and 40, 000 account stubs, producing up to 3 miles of paper tape for computer processing, and ICT 1301 computers currently provide the account processing and document production in this 'turn-around' application.

Error Control

Clearly, in an operation of this size maximum efficiency is of prime importance, and even marginal shortcomings—say less than 1% of entries—can result in a large actual volume of errors. Close control and unending vigilance are therefore maintained at all stages and, since some of the documents processed—the h.p. accounts—are off-set litho reproductions from computer printed masters, the maintenance of this control is no mean task.

The computer system, of which the document reading process is the core, has a number of objectives to fulfil, and in consequence errors can arise for reasons not directly associated with computer working. Nevertheless, all are finally revealed at computer input and, in the normal course of events, the residual error rate remaining at this stage in the system is between 1% and 2%.

These are errors from all causes, but, it should be noted, are after the second processing (or replacement, as the case may be) of documents rejected by the Lector readers at the first pass.

The most recent analysis of these errors was as follows: -

TWO LECTOR READERS—19TH MARCH 1966

No. of documents processed	15, 386		
No. of errors	343		
Percentage of errors	2. 22		
Cause	Number	% of documents processed	% of errors
Omitted from reading process	159	1. 03	46. 36
Clerical or operator errors	84	0. 54	24. 49
Lector machine errors	48	0. 31	13. 99
Printer or off-set litho machines	52	0. 34	15. 16
	343	2. 22	100. 00

This analysis indicates that most of the errors are attributable to human frailty, and that the document readers carry out the translation

of the bar markings into paper tape with remarkable accuracy. The use of the modulus 11 check digit on reference numbers has proved invaluable, as have other precautions such as the standardisation of paper quality for all documents, the provision of folding areas free of bar markings on documents subsequently handled by mailing machines, and the use of orange printing ink on the pre-printed computer bill forms. Orange is the colour to which the Lector readers are least sensitive. With very few exceptions, documents are returned by customers in excellent condition, free from extraneous matter and folded along the original folding areas.

Throughput

Since the Lector machines are hand-fed by female operators, throughput is significant. The system itself is based upon small batches (maximum size 96 documents) each of which undergoes a second pass if necessary—and perhaps replacement of every rejected document—before operations begin on the next batch. This method of working reduces throughput to a maximum of about 1,000 per hour which is undoubtedly less than the rate at which the Lectors would process documents if fed continuously, i.e. if a 'stream' rather than a batch method were possible. The machines are operated by general clerical labour, and fewer operators are engaged than would be needed for a punch/verifying operation of the same magnitude.

Documents

Fig. 68 shows an example of a bill with the stub portion on the right hand side, designed for Lector reading. The first three rows of markings in the upper set comprise a binary-coded decimal equivalent of the account number, and the final two rows represent the amount. The area beneath that allocated for computer-made marks is provided for hand marking of an amount differing from that actually due. Although the amount due is known in the account record, it is provided in a marked form for read-in to facilitate the checking of batch totals.

Conclusions

In common with most other electronic equipment, it is the intermittent fault which gives the greatest trouble on the document readers. A log is maintained of the paper tapes produced by each machine each day, and the volume is such that it is often possible to trace the fault pattern from the computer printed output of errors within the first hour or two of the subsequent day. In fact, these faults seldom occur, which says much for the design and construction of the Lector readers.

The Board is quite satisfied with the system and with the equipment, but is not blind to the advantages which automatic feed and on-line working would offer.

west midlands **GAS** board

PAYMENT SLIP — GAS
FOR OFFICE USE ONLY

Receiving Stamp

F R WELLS & CO LD
2 ALFRED ST
BIRMINGHAM 2

43 22917

Account Reference	
481	007112
Period	Tariff Zone
1/65	1 2
CHARGES	
£	6. 4. 7

METER READINGS	
Present	Previous
4458	4328
GAS USED	
100's of cu. ft.	
150	

YOU CAN PAY this bill at any Gas Showroom,
any branch of the Birmingham Municipal Bank,
or by post to :-

W.M.G.B.
P.O. BOX 12,
SOLIHULL, WARWICKSHIRE.

CHEQUE PAYMENTS - If you pay by cheque
please SEND ONLY THE PAYMENT SLIP. No
receipt will then be issued.

Please cross your cheque and make it
payable to 'W.M.G.B.'

RECEIPT REQUIRED
Enter 'X' here — BILL FORM INTACT
if receipt is not required
with your cheque.

Account billed on 1 APR 20 MAR
Meter read during week ended
Declared Calorific Value 500 BTU. per cu. ft.

AMOUNT DUE
£ 6. 4. 7

481 007112 £ 6. 4. 7
ACCOUNT REFERENCE AMOUNT DUE

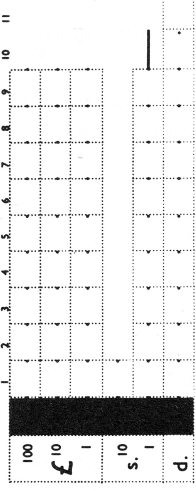


Fig. 68

4. LEGAL AND GENERAL ASSURANCE SOCIETY LTD

The office concerned is a large insurance office, the bulk of whose business is in the Life and Pensions field, and, in consequence, there is a very large Annuity fund. In addition, there is a considerable volume of Fire and Accident business, but this is small in comparison with the Life and Pensions classes. The equipment configuration is an IBM 1401 computer with 16K positions of storage, a 1418 Optical Reader, 2 × 7330 tape units, a card reader and printer.

Reconciliation of Annuity Payments

At the present time the office is despatching about 75,000 Annuity cheques per month. Each annuitant receives from the office a cheque in respect of the net amount due, together with a counterfoil showing the tax deductions. These documents are forwarded in window envelopes, either direct to the annuitant or in some cases to his bank for direct crediting to his bank account. After encashment they are returned via the Insurance Office's bankers in the usual way, together with a detailed statement of account.

The cancelled warrants are kept by the bank in batches, each batch corresponding to one sheet of the statement of account, and are subsequently maintained in these batches within the office. The batches are not sorted in any way, but prior to being passed to the Computer Department a 'bundle marker' is inserted in front of each batch. This is a document of the same length as the warrants themselves, but approximately $\frac{1}{2}$ " deeper in order that it may be picked out by hand later. The bundle markers are identified by the machine by means of a special coding line.

The warrants are fed into the machine still in their original bundles and a punched card containing the batch total is simultaneously fed in. After the data has been read in, examined and re-constructed by program, the totals of each bundle are printed out and agreed with the totals of each page on the bank statement as punched in the cards. A specimen of the counterfoil and warrant is attached—see Fig. 69. The principal difficulty which had to be overcome in this particular job is that, as is well known, it is the custom of banks in this country to cancel cheques by means of a large rubber stamp which may appear anywhere on the document, and it was necessary to ensure that the data was still capable of being read despite this mutilation.

Prior to putting the system into operation a vigorous effort was made to determine whether the banks would be willing to impress their cancellation stamps on the backs of the cheques, as is done in many countries abroad, or at least to standardize the size and position of these stamps. No success was achieved in these efforts, but if operations such as this become more widespread a further approach to the banks would be well worth while.

In order to achieve the best results under present conditions the following steps were taken: -

000 70 :16849 :22

- (a) It was considered that the part of the cheque with the best chance of remaining reasonably unharmed was the strip immediately adjacent to that reserved for the use of the bank at the bottom of the warrant.
- (b) In order to cope with the heavy overprinting by the banks, all the data requiring to be read, namely, annuity number, date of payment, and amount in pence, was written twice in the reading line.
- (c) Each item of data was supplied with its own check digit.

The Optical Reader is quite capable of reading information from almost any part of the document and for an accounting application such as this it would presumably be considered preferable if the information to be read by the Optical Reader were the same as that used by the banks in the normal accounting process. However, this job first became productive in April 1964 and there was a dearth of information regarding the use of the Optical Reader prior to that time, so that bearing in mind the considerations enumerated above, it was felt preferable to provide a special reading line for this purpose rather than attempt to make use of the normal printed data. No difficulty has been experienced since the computer totals are always compared with the totals on the banks statements. For the same reason, although control symbols to distinguish the various fields could have been pre-printed on to the cheque, it was thought at the time that the use of 'lozenges' to break up the data would be preferable in view of the greater flexibility and control over horizontal alignment which might thus be possible.

Considerable numbers of experiments were carried out on blank documents prior to the job becoming productive in order to determine the effect of heavy overprinting in different areas of the cheque. In the course of these experiments it was found that had the reading line not contained duplicated data, a correct reading rate of only 5 per cent was likely to be obtainable. In fact, using duplicated data a reading rate of between 96 and 97 per cent is regularly achieved.

Clearly it is necessary for operators to take considerable care to ensure that the quality of printing is up to standard and, since the documents to be read on any particular run may have been created at any time during the past few weeks, it is vital to ensure that the vertical positioning of the print remains constant from run to run. This is achieved by ensuring that the amount in figures on the form is printed centrally within the 'box' provided.

The job itself is not particularly time critical since it is part of a retrospective accounting reconciliation and it is perfectly reasonable if two or three days' bank clearances are dealt with at one time. The present volume is of the order of 3,000 documents per day.

Recording of Payment of Life Premiums

In this application the policyholder receives a Life Renewal Notice to which is attached a counterfoil which he is requested to return with his payment in a pre-addressed envelope specially designed to hold the counterfoil without folding. (See Fig. 70). Owing to the large number of premiums which are collected in bulk, the volume at present is not in excess of 1,000 documents per day.

KINGSWOOD HOUSE, KINGSWOOD, TADWORTH, SURREY.

KINGSWOOD HOUSE, KINGSWOOD, TADWORTH, SURREY.

LIFE RENEWAL COUNTERFOIL

POLICY NUMBER	00 324596-6
DUE DATE D M Y	10165
CURRENT PREMIUM £	32 16 3

32	16	3
CURRENT PREMIUM £ s. d.		

All

THE CURRENT PREMIUM
WILL BECOME DUE ON
THE DATE STATED AND
SHOULD BE PAID WITHIN
THE DAYS OF GRACE
STIPULATED IN THE POLICY

IF YOUR REMITTANCE IS OTHER THAN AS STATED ABOVE PLEASE SIGN HERE
AND STATE THE REASON:—

PAYMENT MAY BE MADE AS SHOWN ON THE REVERSE OF THE RENEWAL NOTICE.
PLEASE DO NOT MARK IN ANY WAY THE SPACE BELOW THIS LINE

□0003245966□0101659□00000787581□

Fig. 70

In this application, as distinct from the previous one mentioned, the problem of mutilation does not arise to any extent and in consequence the data included on the reading line is not duplicated, so that more time is left available for processing in between reading successive documents. This hope has been justified and at the present time not more than about $\frac{1}{2}$ per cent of the documents are being rejected by the Optical Reader.

For the same reasons as were mentioned in connection with Annuity Warrants, a special line was printed for the purpose of being read optically and it was decided to place this in the same position as had been selected in connection with the Annuity Warrants. In addition, each separate item of data contains its own check digit and also, as the items are not duplicated, there is a further check digit in respect of the other three check digits. The items to be read in fact consist of a Policy No., Due Date and Amount of Premium in pence.

Upon receipt of cash, with attached counterfoils, by the Cashiers, the counterfoils are broken down into batches of suitable size and are again separated by bundle markers and passed to the Optical Reader. In fact four markers are inserted for each bundle. The first bundle marker is sent into the 'normal' pocket and the other three into the 'reject' pocket. All documents in the 'reject' pocket at the end of the pass are then fed in again and this process repeated. By this means any documents rejected on the first or second pass but later accepted are included in the correct batch total for reconciliation purposes. A similar operation was not possible in the case of the annuity warrants as, owing to the length of the reading line, insufficient time was available, during the passing of the document beyond the reading line, for reconstruction and pocket selection.

The vertical alignment of the printing is in this case facilitated by the provision of a fainter inner box at the extreme left of the printing line on the renewal notice and the right of this line on the counterfoil which in the specimen contains the codes 'A11'.

This job is only time critical to the extent that it is obviously desirable that the amount being paid into the bank should be verified as far as possible on the same day.

It is probable that in the future this job will be extended to include recording of payments made by means of standing orders in excess of 3, 000 per day.

Document Quality

It is necessary that all documents to be read by the Optical Reader should be of good quality with a white background. In this case the paper used was Wiggins Teape Sortaspeed Document White Wove of a weight of 90 grams per square metre which is of a quality used by the banks for cheques subject to operation in document sorters.

On the two jobs currently in operation it has been found that the documents are returned to the Insurance Office in very good condition with the exception of the overprinting by the bank's cancellation stamps; efforts to persuade the banks to place their cancellation stamp in a particular part of the cheques being so far to no avail. Returning for a moment to this subject, cancellations performed by perforation as dis-

tinct from overprinting can be particularly offensive since chads from the perforations can adhere to the documents and subsequently transfer themselves to the lens of the Optical Reader. In this connection it has been found that fairly frequent cleaning of the optical reading equipment is necessary to free it from paper dust deriving from the documents.

An attempt has been made to standardize the size of the documents used and they are all $3\frac{1}{2}$ " in depth and approximately $7\frac{1}{2}$ " wide. In consequence we have no experience of feeding intermixed documents of varying size except to the extent that, as mentioned earlier, bundle markers have been designed to protrude beyond the documents themselves, and no difficulties have been encountered in this respect.

Print quality must be carefully maintained if documents are to be read satisfactorily. It is not suggested that this places too severe a burden upon the operators, but undoubtedly reasonable care is necessary, not only as regards the alignment of the characters but also as regards the density. There is a tendency at first to use much too high a print density with the result that the ink from the printer ribbon tends to splash slightly, thus giving the characters a woolly appearance when viewed through a magnifying glass and, in fact, a far higher proportion of acceptable documents is actually obtained when the print density is rather lighter than might at first sight appear satisfactory.

The principal purpose of using Optical Reading equipment has been to avoid the waste of time otherwise necessary if cards had to be punched on receipt by the office of documents which it had itself prepared in the first instance. An alternative would have been for the computer to punch a card, and to pull this from a file at the appropriate time, but card pulling itself can be a very laborious process.

No special type font is used for these documents as the object of the operation was to make use of the normal printed character (UKA3 font). In consequence the usual IBM characters are used for production of the documents on an IBM 1403 printer. It should be noted, of course, that for the purposes mentioned alphabetic characters are not required to be read and any alphabetic characters which might appear, for instance in Policy Numbers, would have to be coded out in the optical reading line. Since, however, this is not a part of the document which has to be read visually, this is unimportant. So far as the choice of Optical Reader is concerned, the IBM 1418 was a natural choice since it was not considered desirable to use mixed equipment supplied by different manufacturers, in view of difficulties which can arise by virtue of divided responsibility.

5. MONTAGUE BURTON LTD

Introduction

The Data Processing Centre of this Multiple Tailoring Organisation is equipped with a Solartron Electronic Reading Automation Machine, for the purpose of reading Till Rolls.

The system, in brief, provides for cash payments, when made at the several hundred branches or at Head Office direct, to be printed by a Register Machine on Till Rolls, which are read by the ERA, and automatically converted to 80 column Punch Cards, which are designed to accommodate six payments in random order. The further processing is carried out by an ICT 1301 computer.

Users of the Credit Facilities offered by the Company are issued with a plastic token, embossed with account number and associated check digit and branch number, making a total of eight digits in all. Payments may be made at any branch. The token is placed in the Register and the embossed data from the token, together with the amount paid, and the balance remaining outstanding is printed onto a Till Roll, in a type font developed by the manufacturers of the ERA machine.

Till Rolls are sent each day to Head Office to be processed on a daily cycle.

The system that was eventually decided upon centred round two main requirements which were considered most important to attain; they were: -

- (a) to eliminate transposition of written characters by branch Cashiers, in particular the account number and check digit.
- (b) speed and accuracy when converting the data to punch cards previously accomplished by hand punch method. Elimination of data verification.

The Documents

Several hundred Till Rolls are processed daily as received from branches.

They may be of any length, the only stipulation being that a 10" head and tail of blank paper is necessary, to switch on and off the scanning circuitry for reading the first and last entries.

The width of the Till Roll is fixed at 100 mm, but facilities are provided for modification should wider paper become necessary.

The use of Till Rolls, was introduced in conjunction with the Reading and System requirements, payments having previously been written by Cashiers on daily cash sheets with all the associated problems of bad figures, bad carbons and transpositions, etc.

The question of paper and carbon quality was important and fully considered in relation to the Reader tolerances, but this presented no problem.

000000000	0000	01003
111111111	1110	99093
222222222	2120	00000
333333333	3133	00000
444444444	4040	00000
555555555	5053	00000
666666666	6166	00000
777777777	7170	00000
888888888	8186	00000
999999999	9199	00000

Fig. 71

A specimen Till Roll showing the character format is shown at Fig. 71. In practice the copy for optical reading is printed in a right to left format i.e. a mirror copy of the one held by Branches, which is in a legible left to right format for normal references.

Character Recognition Equipment

The system used for Character Recognition is optical, based on a flying spot scanner.

The print font was designed by Solartron to give as little stylisation of the characters as possible, yet to facilitate their ease of recognition by the ERA Reader Machine.

The Reader makes an automatic check digit test on the Recognition Output for the first to the fifth and eighth digit in each line of print. Any character or complete row of characters which cannot be machine recognised, after a pre-programmed number of re-scans, may be keyed in via an external key-board by the Operator, and which in turn is also automatically checked on Output.

The line of characters is displayed on an Oscilloscope and any single digit which cannot be read is highlighted for Operator action.

Interfacing

The ERA Machine is coupled to an eighty column Card Punch Unit, which forms part of the system.

The Code Line

The document used as input is a continuous roll and every line is read by the ERA.

Each line consists of 17 digits which are used as follows: -

- (a) digits 1-4 customer account number
- (b) digits 5 check digit
- (c) digits 6-8 branch number at which the account was opened
- (d) digits 9-12 instalment
- (e) digits 13-17 balance outstanding.

The last item (e) is not used within the system and therefore is not read by the Reader.

Printing

Printing is effected by a specially designed Register, and is reproduced onto a blank roll of paper by carbon reproduction. Digits 1-8 are recorded from the embossed characters on the plastic token. The remaining digits (9-17) are printed from embossed hammers which operate when the Cashier sets up the amount of the payment being made.

The usual check for paper alignment, correct loading etc., is taken, this being in the form of Cashier training and Operating Instruction Reminders.

Operational Experience

The system was developed and phased over a period, mainly to allow for the training of Branch staff Cashiers. Processing commenced in June, 1963, and the machine has now been fully operative for some nine months.

The Reader machine requires one Operator only who attends to the processing of Till Rolls, and punch unit requirements. Till Rolls are presented by the Operator, and when the Reader is arrested on a doubtful character(s) the Operator may key-in the queried data as required.

Batches of Till Rolls are pre-sorted according to Branch areas and in days. They are also checked for deficiencies, i.e. short heads and tails, printed data that may be rejected by the ERA machine, e.g., narrow margins.

The document throughput is slowed down somewhat when Till Rolls have few lines of data recorded, as opposed to those that have many. Therefore maximum use of the speed of the Reader may only be attained when Till Rolls of many items are processed in length. In any event however, the output is governed by the speed of the Card Punch Unit, namely 100 cards per minute.

The mis-read error rate experienced to date is better than one in twelve thousand characters. Line rejection occurs when tokens have been damaged and the Reader or Operator cannot make any sense of the check digit routine. It is only on these occasions that the Till Rolls are rejected for special attention.

Line rejections, when they occur, are highlighted by the Reader marking the line of Data as rejected. The Operator has the facility to skip manually to the next line and reading continues until the last entry, following which the Till Roll is rejected to the front of the Reader, as opposed to those fully processed which are deposited in a collecting tray at the back of the machine.

Errors resulting from mis-recognition stem mainly from parts of the character limbs being missing, probably due to damaged tokens for one reason or other.

All data is printed at the time of recording a cash payment; no pre-printed data is used in the system.

The efficiency and speed of operation have had a profound effect on the automatic data processing system now in operation. Staff previously required for resolving account number queries etc., are now used on other duties.

There are on average approximately thirteen and a half thousand items processed per hour. To date approximately 3.9% of available hours is required for down time including scheduled maintenance.

This system utilising techniques for the automatic punching of data is giving very satisfactory results. However like all similar systems the ultimate success depends entirely on the quality of input, which when considered in relation to widely dispersed units and varying staff efficiency, requires continuous attention in order that the necessary standards can be maintained.

6. MERCANTILE CREDIT COMPANY LIMITED

Introduction

The nature of the business is Instalment Finance.

Individual debit accounts are set up for new transactions and maintained on NCR 315 with CRAM.

Daily processing entails: -

- (a) Setting up new accounts.
- (b) Posting cash received and other accounting entries.
- (c) Amending records—changes of address, etc.
- (d) Extracting individual account information for follow up of late payers and defaulters.

All operations are time critical to the extent that: -

- (a) Copies of HP agreements for new transactions must be returned to customer within 7 days of date.
- (b) Follow up notices must be despatched to customer on same day as cash receipts are posted.

Reading accuracy is essential. Control is by check digit, batch pre-lists and daily balancing.

A book of Perf-O-Data coupons is issued for each new account, and is sent to the customer or his Bank (if payments to be made by standing order—present limit of 30 coupons per book). Each coupon is perforated with customer account number and the amount of his monthly payment.

Documents

Not all payments received are accompanied by payment coupons. Many accounts current at present were accepted before Perf-O-Data system was in full use.

7, 500/8, 000 coupons are processed daily.

The general condition of coupons is good but some banks persist in stapling coupons together to the detriment of the reading operation.

5% to 7% of payments received differ from the perforated monthly payment requiring a manual outsort and encoding before reading.

Processing is in mixed size batches but we aim at approximately 150 coupons per batch.

The paper weight is 0.095 g.s.m.

No change was necessary in the document size or paper quality. The previous system used similar forms overprinted with account number by address plate and the paper quality conformed with bank requirements because of possible introduction of MICR.

A sample document is shown at Fig. 72.

ACCOUNT NUMBER	DUE DATE	AMOUNT	£5 & £10 NOTES
0000000000	000000	000000	£1 NOTES
0000000000	000000	000000	10/- NOTES
0000000000	000000	000000	5/- NOTES
0000000000	000000	000000	SILVER
0000000000	000000	000000	NICKEL BRASS
0000000000	000000	000000	COPPER
CREDIT TRANSFER			TOTAL CASH £
CODE NO. 62-00-83			
DATE			
CASHIER'S STAMP AND INITIALS			
To:			
WESTMINSTER BANK LTD.			CHEQUES
BRANCH ACCOUNTS DEPT. L/S			P.O'S. ETC.
COLLECTION ACCOUNT			D £
C			
PAID IN BY			
MERCANTILE CREDIT CO. LTD.			

Fig. 72

Character Recognition Equipment

The Character Recognition system was selected to minimise staff requirements, and to achieve more efficient input. Other forms of data conversion would have required operators and machines, e.g. punching and verifying, add-listing with tape by-product.

The Perf-O-Data system was selected originally because of:-

- (a) Simplicity of book preparation; coupons are gang-punched.
- (b) Legibility of perforated characters; this system was adopted before introduction of EDP.

Interface

The coupon reader punches paper tape. No Interface equipment is required.

Code Line

There are 15 characters in the code line. An end-of-word symbol is emitted by the trailing edge of the coupon. A delete symbol is automatically emitted if the machine fails to read and punch 15 digits. In such cases, the coupon is rejected and can be manually keyed in at the end of batch.

There is no choice in the location of the code line.

Printing of Machine Readable Data

Coupons are made up by our printer in books of 6 and 12. The first coupon in each book is detached after perforating and fed through the Reader to check readability. Paper tape is produced and fed through the computer to check the validity of the serial number against the check digit. Corrections are made before the despatch of the coupon book.

Operational Experience

Perforators installed—July 1962. (For preparation of coupon books.)

Reader installed—February 1964.

Proof Encoder installed—July 1964.

One operator copes with both Proof Encoder and Reader.

Cash received in batches with proof totals. (Payments received without Perf-O-Data coupons are separately batched and processed through an add-lister with tape by-product.)

'Irregular' payments outsourced to back of each batch.

A header coupon containing mode, date and batch number prepared on a Single Document Perforator.

'Irregular' payments perforated by Proof Encoder.

Batch now fed through Reader in two parts, regular followed by irregular, with manual program switch in between.

Output is onto 8 channel paper tape. End-of-batch marker keyed in manually and batches separated by 'tape feed'.

Approximately 8 batches per tape for ease of input.

Nominal speed of Reader is dependent on the coupon length.

Approximately 120 per minute for our coupons. Actual throughput 90/100 coupons per minute.

Approximately 1% rejects. Mainly read by a second run. Failure to read on second run only arises through coupon being perforated out of alignment or mutilated.

Mis-punching from mis-reading seems impossible due to a built in count which automatically rejects short or overlength words.

Subject to initial troubles we have found machine extremely efficient.

Maintenance service monthly. Service very good. Very few breakdowns between service calls due to machine faults. Operator cleans machine weekly.

We have found in practice that if a 'wreck' should occur in the middle of a batch it is simpler to 'delete' and restart the whole batch.

